UNIVERSITY OF OSLO
Department of Informatics

Evaluating the Human Factor in Information Security

Master thesis

Theodoros Nikolakopoulos

Network and System Administration

Oslo University College

Spring 2009

Evaluating the Human Factor in Information Security

Theodoros Nikolakopoulos

Network and System Administration
Oslo University College

Spring 2009

Abstract

Despite the vast research in Information Security, the human factor has been found to lack interest from the research community, with most security research giving focus on the technological component of an Information Technology system. Regardless of any introduced technological solutions, the human factor is still subject to attacks and thus, in need of auditing and addressing any existing vulnerabilities. This research evaluates the human factor by the creation of a survey which examines five distinct user properties. Each of these properties comprise a series of questions, which with their turn assist on confirmation or refutation of five hypotheses. The survey was conducted on two higher academic institutions and distributed to all members of staff who have access on electronic information. Results have shown that the human factor has a significant role in Information Security; it is confirmed that users' behaviour is linked to technology interaction, data importance perception and security oriented education. Furthermore, there is evidence that users who are non vulnerable to various types of attacks, are not necessarily invulnerable to social engineering attacks.

Acknowledgements

I would like to express my sincere appreciation to the following individuals for the support they offered me during the research process. My project supervisor, Siri Fagernes, for her valuable advice and editing; Alva Couch, for sharing with me his research experience by giving strategic advice and editing; Mark Burgess, for his inspirational discussions and guidance; Åsulv Frøysnes, Ole Lycke and Konstantinos Katsifis, for the distribution of the derived survey to Oslo University College and Domi Educational Group respectively; Dag Langmyhr, for his precious help with the survey translation; my colleague, Martin Oppegaard for his assistance with the survey translation and editing; and lastly Maja Oppegaard for verifying the survey translation.

Theodoros Nikolakopoulos

Table of Contents

1 Introduction **1**
- 1.1 Human Error . . . 1
- 1.2 Human Error and System Intrusion . . . 1
  - 1.2.1 Characteristics of a System Intrusion . . . 1
  - 1.2.2 Targeted System Components . . . 2
  - 1.2.3 Insecure by Default . . . 3
- 1.3 Towards Creating a Secure System . . . 3
  - 1.3.1 System Security Threats . . . 3
  - 1.3.2 System Security Goals . . . 4
- 1.4 Security Implementation . . . 4
  - 1.4.1 Linking the Human Factor . . . 5
  - 1.4.2 Problem Statement . . . 5
  - 1.4.3 Targeted Installations . . . 6

2 Literature Review **7**

3 Model and Methodology **13**
- 3.1 Research Method Types . . . 13
  - 3.1.1 Quantitative . . . 13
  - 3.1.2 Qualitative . . . 13
  - 3.1.3 Summary . . . 14
- 3.2 The Research Process . . . 14
  - 3.2.1 Principles . . . 14
  - 3.2.2 User Properties . . . 16
  - 3.2.3 Formulating the Hypotheses . . . 17
  - 3.2.4 Data Collection . . . 19
  - 3.2.5 Data Analysis . . . 22
- 3.3 Human Factors Security Survey . . . 23
  - 3.3.1 Questions . . . 23

4 Results Evaluation **43**
- 4.1 Data Coding . . . 43
  - 4.1.1 Password Scoring System . . . 44
  - 4.1.2 Uncertainty Answers . . . 44
- 4.2 Data Interpretation . . . 44
- 4.3 General Observations . . . 46
- 4.4 First Hypothesis . . . 48

4.5 Second Hypothesis 50
4.6 Third Hypothesis 52
4.7 Fourth Hypothesis 54
4.8 Fifth Hypothesis 56

5 Future Suggestions **59**

6 Conclusion **61**

Appendices **63**

G Informed Consent **65**

H R code for General Observations **67**

I R code for Hypotheses Evaluation **69**

Bibliography **75**

List of Figures

1.1 The Targeted System Components 2
1.2 Linking the Human Factor 5

3.1 The Research Process Flow Diagram 15

4.1 First Hypothesis Results Representation 48
4.2 First Hypothesis x axis Histogram 49
4.3 First Hypothesis y axis Histogram 49
4.4 Second Hypothesis Results Representation 50
4.5 Second Hypothesis x axis Histogram 51
4.6 Second Hypothesis y axis Histogram 51
4.7 Third Hypothesis Results Representation 52
4.8 Third Hypothesis x axis Histogram 53
4.9 Third Hypothesis y axis Histogram 53
4.10 Fourth Hypothesis Results Representation 54
4.11 Fourth Hypothesis x axis Histogram 55
4.12 Fourth Hypothesis y axis Histogram 55
4.13 Fifth Hypothesis Results Representation 56
4.14 Fifth Hypothesis x axis Histogram 57
4.15 Fifth Hypothesis y axis Histogram 57

List of Tables

3.1 The User Properties and Contributing Questions Matrix 18
3.2 The Hypotheses and Contributing Questions Matrix 20

4.1 Functions for the Hypotheses Evaluation 45
4.2 General Observations 47
4.3 Median Values for the axes of First Hypothesis 48
4.4 Median Values for the axes of Second Hypothesis 50
4.5 Median Values for the axes of Third Hypothesis 52
4.6 Median Values for the axes of Fourth Hypothesis 54
4.7 Median Values for the axes of Fifth Hypothesis 56

Chapter 1

Introduction

1.1 Human Error

A system is a collaboration among different entities towards achieving a common goal. An Information Technology system is a system where people and technology, having their own components and related activities, interact for the same purpose. It often occurs that systems fail to function as expected, and likewise for Information Technology systems. It is not only the machine part of the system that must function properly but also the user; when a user fails, human error appears. Human error is actually one of the primary reasons for system failure. This is a result of trying to reduce the human nature into a simplified model. The process of simplifying a model to make it easier to use involves removing elements, which initially might not seem of importance; but the lack of them could have catastrophic results. Such is also the case with Information Technology; systems which are human made, are reduced models of a reality that we want to represent and follow the erroneous patterns of their creators, us. Our analog nature is constrained in a digital world of binary digits which obeys the very same commands that we dictate. Therefore, an error is something that we can expect because there are almost unlimited possibilities for something unexpected to occur.

1.2 Human Error and System Intrusion

1.2.1 Characteristics of a System Intrusion

System vulnerability is a system state that could allow an intrusion. An intrusion is a successful attempt to penetrate a system through exploiting an existing vulnerability. It is not necessary that a vulnerable component will lead to an intrusion; however, a responsible system management should keep vulnerabilities from being exploited. An attacker first tries to identify the weakest point of a system; although solid looking defences might also be exploitable, the weakest point of a system offers a higher attack success rate and therefore it is on average the first to be targeted.[1]

1.2.2 Targeted System Components

The technological part of an Information Technology system consists of hardware, software and data. These components may be targeted directly or not; when attacked directly then the human part of the system is not exploited, when these components are attacked indirectly it implies that the human part of the system is exploited as well. Figure 1.1 shows component interactions.

Figure 1.1: The Targeted System Components

Hardware

Hardware is the physical part of a computer system. It is the lowest level of an Information Technology system and it hosts both the software and data components. Some hardware components are the hard disk, the memory modules and keyboard. Attacks which target the hardware are usually the hardest to defend against as most security measures aim the software and data protection. An example of an indirect hardware-based attack would be a hardware keyboard listener, installed to deceive the legitimate system user.

Software

Software is the part of the computer system which enables all functions by harmoniously managing all hardware resources. Software includes the Operating System, any installed applications and hardware firmware as well; it lies between the user and the hardware and enables the latter to effectively use the former. Most security solutions are software-based, as it is the most targeted system component; an example of an indirect software based attack would be the execution of malicious code on the system through a persuading email attachment. This would allow the attacker to gain system access when the targeted user allows code execution by opening the attachment.

Data

Data is all the system output which is generated by the process of software execution. Data are not directly executed from the hardware but are used in the software layer. As data we could describe a database configuration, the database contents or output of a database query. An indirect attack on data would be to deceive a system user by phone by using a false authentication and request details for a database record. It is worth noticing that unlike the rest of the system components, data cannot be exploited directly but only if

another system component is compromised already. In the given example the human factor exposes the data after his successful exploitation.

1.2.3 Insecure by Default

Even if one has a perfectly designed system technologically, user is still an always possible system point of failure. One could tell that any system relying on human interaction could be insecure by default; considering that all Information Technology systems have the very same intrusion point, human interaction. Furthermore, as the assumption is an idealised reality, the truth is far more insecure than it seems.

1.3 Towards Creating a Secure System

1.3.1 System Security Threats

The first step of creating a secure system in a general context, is to identify the potential threats. The threats of a system could be categorized as interception, interruption, modification and fabrication. These four classes comprise all kinds of threats that a system could encounter.[1]

Interception

The term interception means that information has become available to an external source without appropriate authority. An external source can be a person, a program or a system, and it could be detected or not.[1] Good examples of traced and non traced interceptions could be wiretapping which is not successful and successful respectively.

Interruption

Interruption is when a system component becomes lost, unavailable or unusable[1]. An example would be when the cables connecting a critical system are intentionally destroyed; then system connectivity is interrupted and the resources within automatically become unavailable.

Modification

Unlike interception, modification not only involves an unauthorized party accessing a system component, but also modifying it. Modifications can be detected or not; depending on the technical visibility of changes.[1] An example of detectable modification would be a computer virus which alters the keyboard output; in that way the user will instantly become aware of a system alteration. In the other hand, if the same system is infected by a rootkit instead, although there are changes on the system kernel; the user might not detect any difference in system output or overall experience.

Fabrication

By fabrication we mean the injection of counterfeit objects from an unauthorized party[1]. As these are additional objects it might be easier to detect but it depends on the attackers' proficiency. For example a malicious user could insert a module at a bank database server which would deposit to his account for each transaction a certain very small, and supposedly undetectable, amount.

1.3.2 System Security Goals

Information Security aims to ensure data confidentiality, integrity and availability[1, 2]; three properties which can guarantee data security.

Confidentiality

Confidentiality is present when every system component is accessed by only authorized parties. The term access includes knowing the very existence of the system component, viewing or printing.[1]

Integrity

Integrity ensures that the system components can be modified by only authorized parties or manner. Modification includes writing, changing, changing status, deleting and creating.[1]

Availability

Availability means that system components are accessible to authorized parties at defined times. An antonym of availability would be the *denial of service*, where the access to a particular set of objects is denied at a given time.[1]

1.4 Security Implementation

With all threats mapped, reaching the goal of a secure system would normally be only a matter of cost. However, security is a frequently underrated aspect of technology. Although a responsible process for building an Information Technology system incorporates security, one needs to be aware of common security practices and that the human factor is often the first point of failure. The existence of the security mechanisms does not itself guarantee an *a priori* secure system, just as proper security oriented configuration does not guarantee data protection. For a solid security implementation, human factors should be evaluated, and addressed when necessary.

1.4.1 Linking the Human Factor

Regardless of which technological component[1] gets compromised, a compromise leads to failure of one or more of the security goals. Therefore, the focus of the investigation will be on the potential point of intrusion, which in our subject is the human factor. The attacks will be examined according to human oriented criteria and not technological ones. Figure 1.2 indicates the topology of all the basic components on an Information Technology system under attack and where the human factor is placed within it.

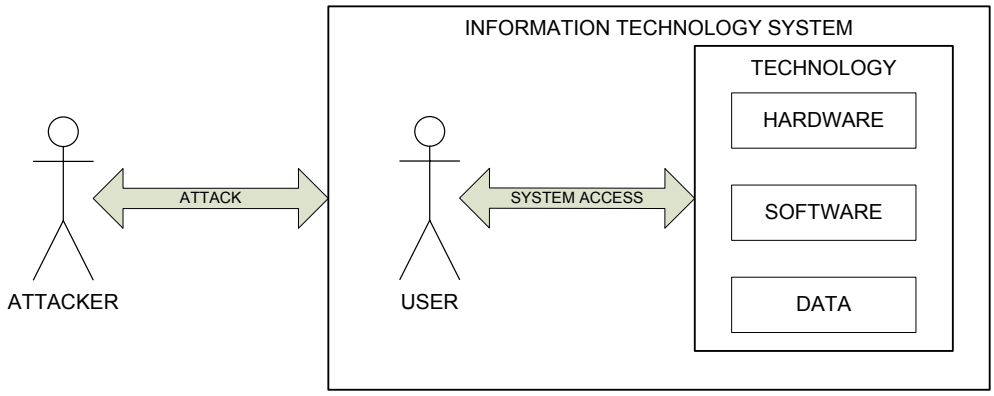


Figure 1.2: Linking the Human Factor

1.4.2 Problem Statement

It is a false assumption that people follow by default secure behavioural patterns and therefore system security expectations should be satisfied. One more false claim is that security is something that can simply be purchased; the human factor can prove the most confident expectations false. The human factor is without doubt a critical point in Information Security. People might take inexperienced decisions which would allow an attacker to take advantage of, or might even intentionally attack their premises; as examining the intentional attack of one on his own premises would involve also topics of wider purport such as psychology, these will not be presently investigated.

The current work will examine the accidental potential points of intrusion from a human factors perspective. There will be an evaluation of the overall existing human factors status. The human factor vulnerability will be tested by developing a survey, which will be distributed to users who interact with the targeted installations. The survey will examine the following user properties:

- Population
- Security Oriented Education

[1] i.e. hardware, software, data

- Security Awareness
- Installation Environment
- Security Policy

These properties will be defined in detail later in the Model and Methodology chapter. The initial suggestion is that people who interact with the examined installations are capable of allowing a system intrusion.

1.4.3 Targeted Installations

The installation environments which will participate on the survey study are two higher academic institutions. The Domi Educational Group from Hellas and Oslo University College from Norway. Both schools have a large number of students; Domi Educational Group has about four thousand students and Oslo University College more than eleven thousand. The number of total employees for each school is also high; the estimated personnel which has access on data is one hundred and fifty for Domi Educational Group and five hundred for Oslo University College. As both clients are academic institutions their assets are joined. Both installation environments are interested into a solid reputation and elimination of any direct or indirect financial losses that a successful attack could involve.

Academic Institutions as Prime Targets

Academic institutions are often targeted as they offer increased anonymity for an attacker. With student accounts being created every semester the population is not static; therefore they accommodate a much easier way for having a very clear idea of the system's internal workings and remaining anonymous at the same time[1]. It has been reported in the past that attackers who are targeting major traffic websites, first practice on university computers[3]. Such institutions are generators of curiosity and scientific challenges which often results in students experimenting upon its own infrastructure[1]. It would be expected therefore, to receive attacks from the inside even without having a malicious intention but just for the sake of proof. An institution is expected to be less secure than a corporate environment and there is usually a less strict maintenance of the system[4]. In addition, an educational institution environment might be less secure for enhancing the student experience with a higher ease of use, something that is also taken advantage of. For all the above reasons, the suitability of an educational institution for examining the human factor vulnerability is excellent.

Chapter 2

Literature Review

This chapter examines the existing literature on the human factor as a potential point of intrusion in an Information Technology environment. Currently there is a lack of research in analysing human factors in Information Security, as the majority of studies is focusing on either usability studies or task analyses[5]. Human factors in relation to Information Technology has been found to be in need of additional interest from the scientific community[6].

A lecture from McCauley-Bell on human factors issues and their impact on Information Security, points out that the increased threats of information technology brought new solutions focused on technological means, while the human factor related work has been extremely limited; with the only notable exception of password generation[7]. Many times organizations overlook the human factor, a factor that security depends upon[8]. Technology is often seen as the immediate answer to Information Security problems[9]. However, despite the fact that many organizations make use of a high number of technical security controls, they still show a non proportional number of security breaches; this happens because Information Security is primarily a human factors problem that remains unaddressed[6]. Since people are the ones who utilize technology, it is just as important to invest in the human factor[9]. A security system, regardless of design and implementation, will have to rely on the human factor; the continuous implementation of technical solutions will fail to handle the people[10]. In addition, Schneier states that technology cannot solve the security problems and believing so shows a lack of understanding of the problems and technology[11]. Mitnik finds technological protection inadequate and argues that users are targeted when the technological attacks fail, in contradiction with most sources that find users targeted first as the weakest system link[12]. Information Security is a set of measures which should be seen as a system and not a single unit[13]. An Information Security system, except of encapsulating the human factor as a component, is also described as a continuously evolving entity[14]. Panko recognizes the intentional threat from both in and out of the organization premises, without analyzing the unintentional exposure of the system to a threat[15]. A security survey from Cisco Systems, revealed that users who work remotely, although they claim to have awareness of security risks, they would still engage into actions which endan-

ger the system security[16].

The unauthorized use of computer systems is made by either accidental or deliberate causes[5]. Accidental causes are any unexpected natural disasters and the human factor[1]; for example, power surges or misconfiguration[5]. The deliberate causes are actions made by conscious choice[2]; for example, using a program flaw to gain access on a computer system[5]. An evaluation of factors which produce security breaches, has shown that sixty five percent of the economic loss in Information Security breaches is due to human error, and only three percent from malicious outsiders[7]. Considering the fact that the efforts to evaluate the human factor in Information Security are basically nonexistent, it is questionable why there has been so much focus on technological means[7].

People as part of the system interact by developing, implementing and using both software and hardware; when a user has poor training, an ideal and flawless software or hardware solution will still not be of any use[13]. Therefore, people will always be a weak system component[13]. Users often perceive their computer systems as a black box, without understanding or the functionality or will to know it[13, 16]. A good example is that users want to operate their computers in the same way as any other household electric appliance[13]. Many users are found to treat confidential information in an irresponsible manner, by having empty passwords or using their name as one; in contradiction to the fact that the same users would never intentionally leave their keys in the outside lock[13]. Regardless of the partial automation that is introduced, people are without doubt involved in technology[6]. Therefore, there is a probability for human error which may result in system exposure[6].

As employees have by *de facto* access and knowledge about the system, they are themselves a potential point of intrusion; therefore, the security of an Information Technology system is greatly affected[17]. Security issues may come into surface when the skills of the employees are higher[17]. This could occur if users would need to use additional software or by considering that they have the knowledge for exploiting any existing system vulnerabilities; additional software could increase the attack surface and an employee with exploitation knowledge could willingly attack the organization from inside[17]. However, it could occur that users with higher technological skills usually require software that they already have the necessary administrative and security configuration skills for. Therefore, this would be something that might not be necessarily true and requires further investigation from the research community.

Security breaches are often caused by careless and unaware users[14]. The majority of people want to get their jobs done more than they are interested in protecting themselves; a behavioural tendency that gives surface for attacks[11]. In addition, most people do not understand subtle threats and they engage into actions which might expose the system[11]. One more view that was not mentioned by any of the related sources is the exception handling, or differently how people might react when something unexpected occurs; many times

[1]The accidental human derived causes can be also called human error

[2]i.e. attacks

attackers rely on the alternative actions that people might take when they encounter something for the first time[11].

Another human factors vulnerability is social engineering; it often happens that attackers directly exploit the user by persuading them to do what they want[11]. Social engineering is a highly effective attack which bypasses every technological protection[11]. Attackers usually are taking advantage of a users trust; building this relation before the attack if necessary[12]. Additionally, quite frequently social engineering attacks rely on the lack of authenticating someone, especially when the correspondence happens through a telephone call; then the attacker might for example pretend to be a person with authority or a fellow employee in need of help[12]. Schultz highlights the unsolved user awareness problem which is confirmed from a mentioned survey, showing that twenty percent of users would not avoid opening email attachments[6].

End users are usually less trained, experienced and security-aware than the Information Technology staff; that makes them the most subject to attack personel in an organization. On top of this, the security vulnerabilities for workstation computer systems are found to be much more than for the server ones; increasing dramatically the possibility of being targeted and the importance of security at workstations. With the workstation accepted as the weakest link of an Information Technology system, an organization can succeed security wise only if it incorporates workstations and their users into the defence frontline.[18]

Security is outlined as a continuous process which require a stopless investment in both technology and users' education; technology, as only a part of it, cannot be the only component for having a secure infrastructure[14]. Users should get educated about risks and responsibilities; education and awareness[1] are identified as key factors in addressing the human element of security[16]. Except giving users an understanding of threats existence, it is also proposed to convince users of the need for security; people would then follow the security requirements in a given situation[1, 16].

A Masters thesis which evaluates Information Technology security performance, also conducts a human factors evaluation according to awareness, training and education[8]; however, the evaluation is made from an organizational perspective and not user-wise, so the results do not come from the users.

A booklet published by a technical group with the topic of system security, is found to be oriented on asset threats and not security threats on assets, so it could be described as a more general reference. The target group of this publication is system administrators and it is written as a sum of good practices and not a scientific publication. The human factor intrusion probability is only categorised under the intentional human threats, while there is a clear possibility of a security incident due to lack of knowledge of the human factor. Although it is noted that the security issue is mainly a problem born from people; the suggested addressing is mostly software based, attempting to solve a human oriented issue by altering the technological component of the system. The authors suggested user-centric solution by enhancing education, is made only in regards to social engineering.[19] Finally, there is an insufficient identifica-

tion of the human factor as an unintentional threat and a narrow suggestion of using user education for preventing attacks on the system.

A publication of a similarly general perspective which aims to cover security in an enterprise environment, was also found to lack in consideration of the human factor role in security[20]. Although the fact that the threat sources are identified to originate from both inside and outside the premises, in the given defence examples, the threats are perceived to be out of premises, always with the intention and attacking through technical means. Therefore, similarly as the previously reviewed source, the majority of the defence mechanisms are either software or hardware oriented. A very brief reference on social engineering is again made with suggestions for reducing such occurrences. The likehood of an accidental exposure of the system through a user is not evaluated; however, in the appendices the section of the Enterprise Systems Security Review does include a human factors checklist which mentions education and awareness, as two attributes which should be examined.

Panko suggests that for designing or auditing security operations, the principle of having clear roles should be implemented[15]. The roles define who does what and determine procedures[15]. Another interesting contribution is the suggestion to do user training[15]; compiled by three parts, security awareness, accountability and self-defence[15]. The security awareness training would aim to help users understand the existence of threats by utilizing attack patterns and case studies[15]. Accountability training would help users be familiar with actions that they should either avoid or not, according to specific rules and the understanding beneath them[15]. By the term or self-defence training, it is meant to prepare a user for taking an appropriate action during an attack; in addition, part of the self-defence training should be using the users on detecting problems or reporting improper user behaviour[15]. An alternative solution of Mitnik is having basic training for everyone and further training according to the users' specific position[12]; a solution which takes into consideration the role principle that Panko gives.

Hinson recognizes awareness as the most cost-effective security control and makes a suggestion on how to optimise control investment; however, without giving a solution[9]. In regards to proactive risk management, it is found that many organizations evaluate new products and do periodic testing on the technological part of their systems, but very few make a serious attempt to identify risks in relation with the system users[9]. Improvements in security require improved understanding of feedback[10]; since the feedback from the technological component can be automated[3], a remaining uncovered topic is the one of human factors.

While Kraemer creates a human factors evaluation method for computer and Information Security, it does have certain constraints. The derived vulnerability evaluation follows a technical vulnerability audit and it takes place on top of vulnerabilities with the human factor components[5]; this limits the possibility of having a human factors vulnerability evaluation without inspecting the technological component. In addition, the vulnerability evaluation is

[3] i.e. software/hardware monitoring

made only according to the earlier found technical vulnerabilities[5] and therefore there is a large part of the human factor, the non technical unintentional vulnerabilities, which remain unaddressed. The feedback comes through qualitative interviews with the involved network administrators and not the end users[5]; while by doing the opposite the obtained information would be more prosperous and realistic. Furthermore, the evaluation is qualitative and based on results which come from a qualitative analysis software package[5]; which raises a risk of inconsistency as the results may vary if an improper categorization is made.

The previously investigated sources constitute the scientific basis for the exploration of human factors as a potential point of intrusion. The human factor is without doubt a critical part of Information Security[6, 7, 8, 9, 10, 11, 13, 14, 16, 17, 19]. This occurs as users expect technology to be there and work for them[13, 16] or simply because they overtake any security mechanism for getting their jobs done[11]. With users being the most targeted and vulnerable link of an Information Technology system, their defence should be of a higher priority[18]. The primary focus of an investigation should be on the education and security awareness of the user[9, 16, 19, 20], two attributes which may define exploitability as well. For addressing the human factor vulnerability, education is recognized as a key factor[1, 9, 11, 12, 14, 15, 16].

Chapter 3

Model and Methodology

This chapter goes through the approach used in the current study for evaluating the human factor in the context of Information Security. The potential of an unintentional security exposure of the system is examined. This approach contributed to exploring the human factor in relation to attributes which affect behaviour. The entire work is kept as general as possible[1] in order to enable a later use in future research.

3.1 Research Method Types

There are two main research methods, quantitative and qualitative. The distinction between quantitative and qualitative methods is made in accordance to the question which is asked, the method behind the answer, and the precision that one requires[21].

3.1.1 Quantitative

Quantitative research applies mathematical modelling and connects the product of research with it. In a quantitative method, the outcome will be measured in relation to a quantity. This method is used in research where measurables can be enumerated and mathematical relationships are known. A weakness of quantitative research is when it comes in need to investigate many, different realities in various depths[22]. The greatest advantage is a more concrete framework and that the data are in an easier to analyse form[22]. During the early steps of evaluating research methodologies, quantitative research looked like the most favoured choice. However, as the subject of study is anthropocentric and quantitative research makes it more difficult to receive a broader spectrum of answers, there was an evaluation of the qualitative research as well.

3.1.2 Qualitative

Qualitative research involves measuring data which is usually related to human actions and the grounds behind them. Qualitative research is mostly used

[1] i.e. not constrained to examine academic institutions only

in behavioural sciences[21]. Qualitative data cannot be always quantified and measured in relation to a quantity; thus, qualitative research is inefficient when it comes to identifying, measuring or quantifying a single statistic[22]. However, an advantage of this research method is the ability to examine given phenomena with respect to multiple human perspectives. The free nature of research allows a more rich input that might contribute to a more specific learning outcome[22]. Qualitative research is more appropriate for human oriented study research. Lack of numeric scoring allows freedom of choice on both questions and answers, and can offer a great input of knowledge to the study. A great disadvantage in comparison with quantitative research however, is that the data cannot always be quantified.

3.1.3 Summary

Quantitative methods, although lacking the flexibility of examining multiple perspectives, makes it possible to examine most areas of interest through carefully predefined questions. The capability to analyse the data without engaging into a process of quantification makes the entire process faster; which would result on the ability of having a larger target group as well. While qualitative methods offer significant freedom of input, it also limits the number of people who can participate in the evaluation; interviewing several people and analysing the results can be very time consuming. In addition, it can be difficult, if not impossible, to conduct a remote interview at the participating institute in Hellas; an email interview correspondence could be a lengthy process, in which not many people will be willing to engage. As the human factor evaluation outcome may vary from user to user, having a smaller data pool could result in a less pragmatic result; for this and the rest of above stated reasons, the desired outcome is best addressed by following a quantitative research method.

3.2 The Research Process

For the examination of the human factor, there were defined user properties, highlighted by the reviewed literature or considered to affect the human behaviour in an Information Technology environment. These user properties are investigated via corresponding questions. Many of these questions contribute to defining several user properties. The next step was to create hypotheses to be tested for various user properties. At the final stage, the data collection and analysis takes place. The followed research process flow is visualised in Figure 3.1.

3.2.1 Principles

The validity of this research depends upon the presence of prerequisite established and consistent knowledge. The first used scientific principle states that

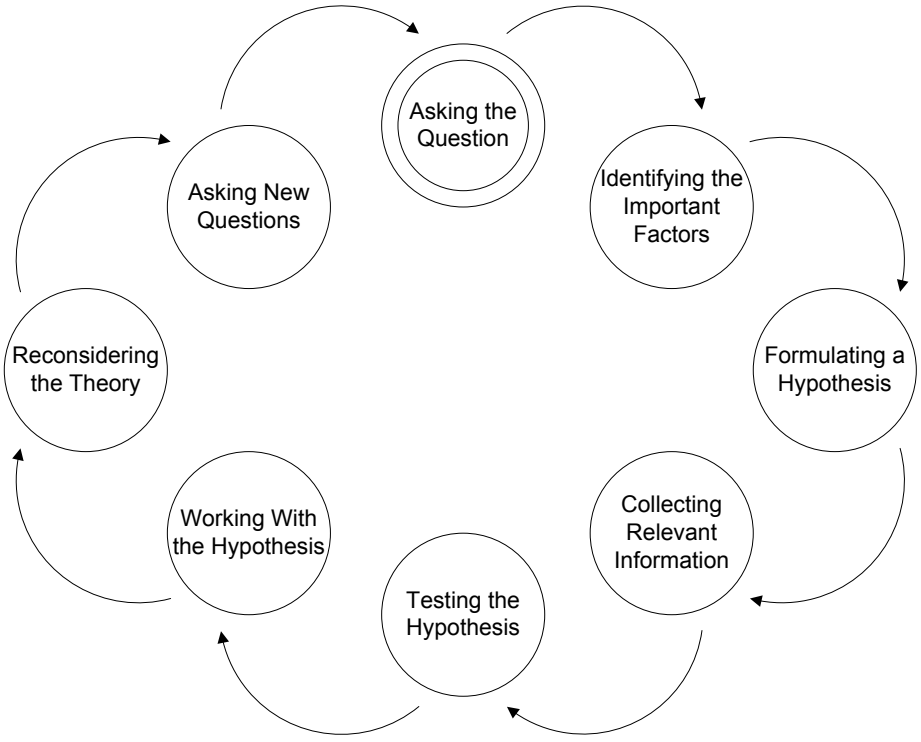


Figure 3.1: The Research Process Flow Diagram

"Security mechanisms do not result on an *a priori* secure system"; it has been shown that regardless of the introduced security mechanisms on a system, they would still be insufficient to ensure security. The second used scientific principle states that "Man as part of the system can be a point of intrusion"; it is proved that man can have a significant role on the system security, as a factor that interacts with the system it can always be a point of intrusion. If these two principles would not be considered as established, the current work would not apply in the given circumstances.

3.2.2 User Properties

The primary aim of defining properties for the human factor was to cover different perspectives. In the course of reviewing the existing work on identifying and addressing the human element of security, education and awareness were found to be repeatedly mentioned among the sources as key factors[9, 16, 19, 20]. However, there are more elements which would need to be examined for more extensive coverage; in addition to *security oriented education* and *security awareness* the user properties which were added were the *population*, the *installation environment* and the *security policy*. All the defined user properties may affect the behaviour of a user in the context of an Information Technology installation.

For investigating user properties, several questions were made in relation to each; many of these questions contribute knowledge on more than one user property, they were categorized on the one to which they contribute the most and also linked to related ones. Not all derived questions give value to a hypothesis, but all give value to at least one user property.

Two additional user properties of *responsibility* and *compliance* were found to arise from a users behaviour but not to define it. For this reason they were included and linked to the questions which would contribute knowledge, but no questions were categorized as specifically measuring them, nor they were used for the evaluation of any of the hypotheses. The user properties along with the contributing questions can be seen as a relational matrix in Table 3.1; where the user properties are listed horizontally and the questions[2] vertically.

Population

The user property of *population* examines individual user characteristics. The questions under this category give an insight on the users experience and interactivity with technology, data protection responsibility and security incident record. The addition of this user property assists in examining the results per certain population criteria.

Security Oriented Education

The user property of *security oriented education* helps to evaluate the knowledge level of a user. This investigates if previous security training exists and the

[2]Categorized by user properties

knowledge of a user about password practices[3]. Apart from the highlighted importance which this property was found to have in Information Security[9, 16, 19, 20], its inclusion allows the evaluation of the human factor by education criteria.

Security Awareness

The user property of *security awareness* evaluates the user's consciousness of various points of intrusion; this takes place by giving cases which could potentially lead to a system compromise. This could also be considered as a vulnerability evaluation of each user. While *security oriented education* examines practical and basic knowledge, *security awareness* examines the existence of theoretical knowledge which would help a user to identify, and therefore avoid, a potential threat. *Security awareness* has been identified as a major element of the human factor in Information Security[9, 16, 19, 20].

Installation Environment

The user property of *installation environment* investigates the user's security alertness and consciousness in relation to the work environment and the environment itself. An installation environment may for example restrict the user's behaviour for reducing the possibility of a system exposure; or oppositely, it could allow such user's freedoms which could lead to system vulnerability.

Security Policy

The user property of *security policy* questions examine the installation environment policy. This user property examines cases in which the policy can control the human factor and thus prevent exposure. While *security policy* might be enforced through the *installation environment*, policy can also be a directive which relies on a user's will not to be overruled.

3.2.3 Formulating the Hypotheses

The developed questions and the user properties to which they contribute knowledge, in turn support or refute five hypotheses which are made in regards to the human factor. The hypotheses along with the contributing questions can be seen in a relational matrix in Table 3.2; where the hypotheses are listed horizontally and the questions[4] vertically.

[3] i.e. password generation and renewal
[4] Categorized by user properties

			Population	Security Oriented Education	Security Awareness	Installation Environment	Security Policy	Responsibility	Compliance
User Properties	Population	q1	X						
		q2	X						
		q3	X						
		q4	X						
		q5	X						
		q6	X						
		q7	X					X	X
		q8	X					X	X
		q9	X					X	
		q10	X					X	
	Security Oriented Education	q11		X					
		q12		X				X	
		q13		X				X	
		q14		X				X	
	Security Awareness	q15			X			X	
		q16			X			X	
		q17			X			X	
		q18			X			X	X
		q19			X			X	X
		q20			X			X	X
		q21			X			X	X
		q22			X			X	
		q23			X			X	
		q24			X			X	
		q25		X	X				
		q26		X	X				
		q27			X			X	
		q28			X			X	
		q29			X			X	
	Installation Environment	q30			X	X		X	X
		q31			X	X		X	X
		q32				X			
		q33				X			
		q34			X	X		X	X
		q35			X	X		X	X
		q36				X			
		q37			X	X		X	
		q38			X	X			
	Security Policy	q39					X		
		q40			X		X		
		q41					X		

Table 3.1: The User Properties and Contributing Questions Matrix

First Hypothesis

The first hypothesis is that "Security awareness increases for the population with higher interaction with technology". The dimensions which are evaluated are security awareness against interaction with technology. This hypothesis is based on the assumption that people who interact more with technology have the potential to be more security-aware.

Second Hypothesis

The second hypothesis is that "Population who perceives personal and work data as more important to protect would be more invulnerable to attacks". The dimensions which are evaluated are data importance perception against invulnerability. This hypothesis is based on the assumption that people who perceive their data less important to protect are more likely to engage into actions which could expose the system to a threat.

Third Hypothesis

The third hypothesis is that "The population with a higher security education will be less vulnerable". The dimensions which are evaluated are security education against vulnerability. This hypothesis is based on the common-sense observation that people with higher security education have the potential to be less subject to attacks.

Fourth Hypothesis

The fourth hypothesis is that "The population who experienced a security incident is more security-aware". The dimensions which are evaluated are security incident status against security awareness. This hypothesis is based on the common-sense notion that people who had a security incident in the past have the potential to be more security-aware.

Fifth Hypothesis

The fifth hypothesis is that "Invulnerability to social engineering does not correlate with other invulnerabilities". The dimensions which are evaluated are social engineering invulnerability against other invulnerabilities. This hypothesis is based on the assumption that social engineering invulnerability does not correlate with other invulnerabilities due to the difficult to detect nature of social engineering attacks.

3.2.4 Data Collection

The data collection took place in the form of a survey. The derived survey was distributed by email through the Information Technology department of

			$H1^{x,y}$	$H2^{x,y}$	$H3^{x,y}$	$H4^{x,y}$	$H5^{x,y}$
User Properties	Population	q1					
		q2					
		q3	Y				
		q4	Y				
		q5	Y				
		q6	Y				
		q7		X			
		q8		X			
		q9				X	
		q10				X	
	Security Oriented Education	q11			X		
		q12			X		
		q13			X		
		q14			X		
	Security Awareness	q15	X	Y	Y	Y	Y
		q16	X	Y	Y	Y	Y
		q17	X	Y	Y	Y	Y
		q18	X	Y	Y	Y	Y
		q19	X	Y	Y	Y	Y
		q20	X	Y	Y	Y	X
		q21	X	Y	Y	Y	X
		q22	X	Y	Y	Y	X
		q23	X	Y	Y	Y	X
		q24	X	Y	Y	Y	X
		q25	X	Y	Y	Y	Y
		q26	X	Y	Y	Y	Y
		q27	X	Y	Y	Y	X
		q28	X	Y	Y	Y	X
		q29	X	Y	Y	Y	X
	Installation Environment	q30	X	Y	Y	Y	Y
		q31	X	Y	Y	Y	Y
		q32					
		q33					Y
		q34	X	Y	Y	Y	X
		q35	X	Y	Y	Y	X
		q36					
		q37	X			Y	
		q38	X			Y	Y
	Security Policy	q39					
		q40	X			Y	Y
		q41					

Table 3.2: The Hypotheses and Contributing Questions Matrix

each participating institution at the users' work email addresses. The users must visit a link, complete the survey and submit electronically their answers. As various data fields are essential for the overall evaluation of a user, all questions were made compulsory. Allowing incomplete entries could result in more results to analyse, but it would add significant complexity to data analysis.

The survey was submitted to all members of staff who have access to query or edit electronic information. Any other member of staff would be considered irrelevant as the survey is focused upon people who have access on an Information Technology system. The authorization for conducting the survey was granted by the Information Technology department of the participating institutions. The survey was distributed to all the target population and to increase potential participation from the educational institution in Norway, it was translated to Norwegian language prior to distribution.

As the component under investigation is the user, the results might be affected by a various environmental factors such as cultural or social. The institutions where the experimentation will take place are heterogeneous environments, with people of various technical expertises, who interact with an Information Technology installation. The participating institutions are known to differ in the number of employees, number of enrolled students, and geographical location. Another factor that might give erroneous results, is that people could reply differently than they would act; for example users could reply that they would not enter their credentials on a website which starts with "http://", but in real life they would not check if this is the default for their web based email service.

Software Tools

The software tool which was used for creating the survey and collecting the data was Google Docs[5]. Google Docs is a web based application that allows the creation of various document types. Among the documents that can be created, there are forms which function as an interface for inserting data to a spreadsheet. This represented great time savings over the option to distribute the survey in a printed form, and thus it was the preferred method for data collection. A disadvantage of Google Docs is that it is not possible to prevent people from submitting multiple replies; something which could result in duplicate entries.

Ethical Considerations

For informing the survey participant of the followed ethical practices, a page of informed consent was added on the very beginning of the survey. This page included:

- The purpose of the research

[5]https://docs.google.com/

- The name and contact details of the conductor
- The identification of the authorization contact at the given institution
- An assurance of non-coercion
- An assurance of confidentiality
- An assurance of privacy
- An assurance of protecting from harm and offer to withdraw from completing the survey if this would be considered harmful

Conforming to the basic principles of ethical research, the only collected data is the participants' answers and the timestamp of submission. The timestamp value was only used to remove accidentally submitted duplicate records. The survey took place for the duration of a week for each institution, after which it was not available anymore online. The full text of the informed consent page can be found in the Appendices.

3.2.5 Data Analysis

The data analysis evaluated whether the hypotheses are verified or not, and drew general conclusions in regard to the users' answers. The fact that users from two different institutions participate, can enhance the results validity by comparing them to each other. Therefore, other than analysing the results individually, a comparison was done as well. Furthermore, there are several questions which contribute to user properties and hypotheses; this enhances the factor analysis which will be used, by having many questions contributing to a factor score[21, 23]. The aim of using factor analysis in results interpretation, is to have one factor for each measurable dimension per hypothesis.

An important point which could influence the results is that of relating the questions with a hypothesis dimension and the lack of using weighting factors. If the questions which are linked to a factor will not contribute significantly or are wrongly assigned to this factor, then the results will be less accurate or even invalid. One more issue is that several of the examined attack cases, are less likely to occur; for example, an email with a malicious attachment is a more frequently occurring attack than someone visiting a user's office and installing a keyboard capture device. During question creation, the weighting factors of each question were not taken into consideration as this would be a lengthy process, requiring a comprehensive analysis of the importance of each.

Software Tools

The data analysis was performed by using the R[6] software for statistical computation. As Google Docs exported the data in a spreadsheet format, the conversion to a readable by R format was made by Microsoft Office Excel[7]. Before

[6] 2.9.0 release
[7] 2007 SP2 release

exporting the data, to reduce sampling error, all duplicate records with timestamp less than a minute were removed. The developed R script which was utilized for analyzing the results can be found in the Appendices.

3.3 Human Factors Security Survey

3.3.1 Questions

This section presents the derived questions for measuring each chosen property. All questions are designed to have predefined answers, for which the survey participant has to choose one or more options[8]. For each question there are listed:

- The question measurable
- The actual question[9]
- The predefined answer options[10]
- The user properties and the hypotheses[11] that the question might aid in testing
- The expected information that the question may supply

Population

The population property examines personal characteristics of each individual that completes the survey. Such are the experience and interactivity with technology, the data protection responsibility and the security incident record. By including this user property later there can be an evaluation of the results per certain population criteria.

First Question The first question requires from the user to state his level of experience on using a computer system. The question is:

What is the level of your computer skills?

The user has to choose one of the following options:

- Beginner (e.g. word processing, Internet browsing)
- Intermediate (e.g. installing programs, installing devices)
- Advanced (e.g. troubleshooting and administrative tasks)
- Professional (e.g. administering systems for other users)

This question contributes to the user property of *population*. It is not used in the evaluation of any hypothesis but it adds important knowledge regarding the population.

[8] The twelfth question is the only exception
[9] With the relevant help text when existed
[10] With the relevant help text when existed
[11] If any

Second Question The second question requires from the user to choose the number of years that he is using a computer system. The question is:

How many years of experience in working with a computer system do you have?

The user has to choose one of the following options:

- Less than a year
- 1-5 years
- More than 5 years

This question contributes to the user property of *population*. It is not used in the evaluation of any hypothesis but it adds important knowledge regarding the population.

Third Question This question measures the user interactivity of the user with technology, by measuring the Internet usage in hours per day. The question is:

On average, how many hours per day do you use the Internet?

The user has to choose one of the following options:

- Less than 3 hours
- 3-8 hours
- More than 8 hours

This question contributes to the user property of *population*. It is used at the first hypothesis as it directly gives an insight into user interactivity with technology. It may be assumed that a user who browses the Internet more hours per day, has already adopted technology to a greater extent than one who does not.

Fourth Question This question measures the user interactivity of the user with technology, by measuring the frequency with which a user checks his email. The question is:

On average, how often do you check your email?

The user has to choose one of the following options:

- At least once per hour
- At least once per day
- At least once per week
- Less than once per week

This question contributes to the user property of *population*. It is used at the first hypothesis for measuring the user interactivity with technology. A user who checks email more often, is considered to have a more frequent contact with technology than one who does not.

Fifth Question This question evaluates if the targeted population uses instant messaging applications. The question is:

Do you use any instant messaging applications?

(e.g. Google Talk, Skype, Windows Live Messenger)

The user has to choose one of the following options:

- Yes
- No

This question contributes to the user property of *population*. It is used in studying the first hypothesis as it gives a view of the users' interactivity with technology. Population which uses instant messaging applications has one more way of interacting with technology, in contradiction with population who does not.

Sixth Question This question evaluates if the targeted population is a member of a social networking platform. The question is:

Do you have a profile on any social networking web site?

(e.g. Facebook, MySpace, Windows Live Spaces)

The user has to choose one of the following options:

- Yes
- No

This question contributes to the user property of *population*. It is used in studying the first hypothesis as it gives an insight of the user's exposure to technology. A user who is member of a social networking platform, could have a higher interaction with technology than someone who is not.

Seventh Question This question evaluates the population feeling of responsibility when it comes to personal data protection. The question is:

How important is it to protect your personal data?

The user has to choose an option from the following range:

- 1. Extremely important
- 2.
- 3.
- 4.
- 5. Not important at all

This question contributes to the user properties of *population*, *responsibility* and *compliance*. It is used at the second hypothesis as it shows the population responsibility at personal data protection.

Eighth Question This question evaluates the population feeling of responsibility when it comes to work data protection. The question is:

How important is it to protect your work data?

The user has to choose an option from the following range:

- 1. Extremely important
- 2.
- 3.
- 4.
- 5. Not important at all

This question contributes to the user properties of *population*, *responsibility* and *compliance*. It is used to study the second hypothesis as it shows the population responsibility at work data. In combination with the previous question it can also add an insight on the different importance level that people might have between home and work data protection.

Ninth Question This question queries the online fraud incident history of the population. The question is:

Have you ever been victim of online fraud?

(e.g. identity theft, phishing)

The user has to choose one of the following options:

- Yes
- No
- I don't know

This question contributes to the user properties of *population* and *responsibility*. It is used at the fourth hypothesis as it reveals whether a certain type of security incident did exist.

Tenth Question This question queries the population security history when it comes to malicious software. The question is:

Has your work or home computer ever been infected by malicious software?

(e.g. keylogger, rootkit, spyware, virus)

The user has to choose one of the following options:

- Yes

- No
- I don't know

This question contributes to the user properties of *population* and *responsibility*. It is used at the fourth hypothesis as it reveals whether a certain type of security incident did exist.

Security Oriented Education

The user property of security oriented education investigates if previous security training exists and the knowledge of a user about password selection. This sector contributes significantly to study of the third hypothesis where population with a higher security education is evaluated.

Eleventh Question This question queries whether the user has ever received computer security training. The question is:

Have you ever received computer security training?

(e.g. lecture, presentation, seminar)

The user has to choose one of the following options:

- Yes
- No

This question contributes to the user property of *security oriented education*. It is used at the third hypothesis as it is expected to give a view of the population who has security training.

Twelfth Question This question tests which sources are used from the population for creating a password. The question is:

Which of the following do you use for generating your passwords?

Check all boxes that apply

The user has to choose one or more from the following options:

- Personal Information (e.g. date of birth, place of birth, address, name)
- Dictionary words (e.g. apple, backyard, cloud, door)
- Phrases (e.g. It is not in the stars to hold our destiny but in ourselves)
- Numbers (e.g. 0, 1, 2, 3)
- Symbols (e.g. !, #, %, &)
- Lowercase letters (e.g. w, x, y, z)

- Uppercase letters (e.g. A, B, C, D)

This question contributes to the user properties of *security oriented education* and *responsibility*. It is used to study the third hypothesis because it shows the knowledge of users about creating strong passwords. A good password on a healthy system prevents unauthorized access to external sources.

Thirteenth Question This question tests the users preference on password length. The question is:

On average, how long in characters are your generated passwords?

The user has to choose one of the following options:

- Up to 7 characters
- 8 or more characters
- As short as the system accepts

This question contributes to the user properties of *security oriented education* and *responsibility*. It is used to study the third hypothesis as the knowledge of password length is part of the knowledge for a strong password. It is not only enough to have a well generated password; the longer a password is, the harder it would be for a malicious user to recover it.

Fourteenth Question This question measures the frequency with which a user changes his email password. The question is:

On average, how often do you change your personal email password?

The user has to choose one of the following options:

- At least once per 3 months
- At least once per 6 months
- At least once per 12 months
- Only when required by the system
- Never

This question contributes to the user properties of *security oriented education* and *responsibility*. It is used at the third hypothesis as it shows a users knowledge on a secure password policy. A frequently changed password is harder to be recovered than a less frequent of the same strength. The change of a user's personal email password is not enforced by any policy other but himself; therefore, it shows one's responsibility when it comes to personal data security.

Security Awareness

The user property of security awareness evaluates the user's personal security alertness for various points of intrusion. It is the only property which is used in the study of all five hypotheses. The social engineering questions are included in this property because, although such attacks might take place at work environment, their success relies on users' lack of awareness.

Fifteenth Question This question checks if the user has the same password for many different accounts. The question is:

Do you reuse the same password for several user accounts?

(e.g. personal email account, work email account)

The user has to choose one of the following options:

- Yes
- No

This question contributes to the user properties of *security awareness* and *responsibility*. If someone uses the same password for more than one account, then in case that an attacker gains access to one of them, the security of all might be at risk[24]. It is also often that people might have one strong password which they use for several user accounts; thus making them all vulnerable upon one's compromise, regardless of the password strength.

Sixteenth Question This question checks if the user would write down a complicated password. The question is:

If you were not able to change a password that is difficult to remember, would you write it down?

The user has to choose one of the following options:

- Yes
- No
- Maybe

This question contributes to the user properties of *security awareness* and *responsibility*. By writing down a password which is difficult to remember then it is much easier for an attacker to get access. One common example of such bad practice is when users are writing their passwords and attach them on either their computer system or desk; then once an attacker can be where the computer system is, the very purpose of having a password at all becomes obsolete.

Seventeenth Question This question checks if the user is visually securing his credentials entry. The question is:

Do you prevent others from watching you type when you enter your username and password?

The user has to choose one of the following options:

- Yes
- No

This question contributes to the user properties of *security awareness* and *responsibility*. The lowest technologically form of eavesdropping is when someone captures a password by simply watching a user entering it[24]. When a user enters his username and password, it should be of his responsibility to not allow anyone to look at his typing. Even a partial recognition of the password might significantly accelerate the ability of an attacker to recover the whole.

Eightteenth Question This question checks if the user needs a password to login to his home computer. The question is:

Do you use a password to login to your home computer?

The user has to choose one of the following options:

- Yes
- No

This question contributes to the user properties of *security awareness*, *responsibility* and *compliance*. When a user has a password enabled account at his home computer he prevents a wide spectrum of attacks. A password enabled home computer, could furthermore show the responsibility that one has towards a simple security practice, over the convenience of not using a password to login.

Nineteenth Question This question checks if the user has installed antivirus software on his home computer. The question is:

Do you have antivirus software installed on your home computer?

The user has to choose one of the following options:

- Yes
- No
- I don't know

This question contributes to the user properties of *security awareness*, *responsibility* and *compliance*. The antivirus software is an essential part of a system security that prevents the execution of malicious code. A responsible user should make sure that antivirus software is installed on his home computer. Although the user's home computer might be of such an architecture or having installed such an operating system that is perceived more secure, there are several threats in the wild for all architectures and operating systems and in general principle although there is still a significantly lower possibility of having certain systems infected, the possibility still exists[25, 26, 27, 28, 29, 30].

Twentieth Question This question evaluates if the user would allow someone to use his home computer under his supervision. The question is:

Would you allow someone to use your home computer with your supervision?

The user has to choose one of the following options:

- Yes
- No

This question contributes to the user properties of *security awareness*, *responsibility* and *compliance*. The physical access on a computer system usually tears down most of the defensive mechanisms against malicious users. Furthermore, an attacker might execute tasks in such a way, which would not be suspicious at all in the eyes of the supervisor. However, a supervisor with professional computer knowledge could efficiently monitor every activity that one engages when using his computer system.

Twenty-first Question This question evaluates if the user would allow someone unsupervised to use his home computer. The question is:

Would you allow someone to use your home computer without your supervision?

The user has to choose one of the following options:

- Yes
- No

This question contributes to the user properties of *security awareness*, *responsibility* and *compliance*. Although the physical access might lead to a security compromise, when the actions of an attacker are supervised, there is a limitation and absolute dependency on the supervisor skills to recognise any malicious actions. If there is a total lack of supervision, the attacker is free to act and this would almost guarantee a system compromise.

Twenty-second Question This question evaluates if the user would open an email link or attachment from a familiar email address. The question is:

Would you open an email link or attachment from an email address you recognize?

The user has to choose one of the following options:

- Yes
- No

This question contributes to the user properties of *security awareness* and *responsibility*. Specially crafted email links or attachments, might deceive the user to enter his credentials or allow code execution that would compromise the system security. An email address which seems familiar would be less suspicious for the receiving end. In addition, not all users have the technical expertise to recognise spoofed email headers and from the ones who can, there is few who validate in depth the origin of an email.

Twenty-third Question This question evaluates if the user would open an email link or attachment from an unfamiliar email address. The question is:

Would you open an email link or attachment from an email address you do not recognize?

The user has to choose one of the following options:

- Yes
- No

This question contributes to the user properties of *security awareness* and *responsibility*. Despite the fact that a user might not recognise the email address that send him an email, it might be of such a content that might temp him to open an email link or attachment; an action which could trigger a system intrusion.

Twenty-fourth Question This question evaluates if a user would share his account credentials with someone else. The question is:

Would you share your username and password with someone else?

(e.g. friend, colleague, assistant, teammate)

The user has to choose one of the following options:

- Yes
- No

This question contributes to the user properties of *security awareness* and *responsibility*. One of the less expecting but yet often occurring ways of getting access to a system is to ask someone for his username and password. Many people would even offer their account credentials by themselves in order to get help or, for the sake of offering comfort, to a usually familiar person[24]. The security risk of such an action is very high as there is not only the potential misuse of the credentials to be considered, but also the overall system security responsibility and risk, which expand to one more person.

Twenty-fifth Question This question checks if the user would enter his account credentials on a website which uses the insecure Http protocol. The question is:

Would you enter your username and password on a web site whose address starts with "http://"?

(e.g. http://www.example.com/)

The user has to choose one of the following options:

- Yes
- No

This question contributes to the user properties of *security oriented education* and *security awareness*. The Http protocol transfers the data in a plain text format, which allows an attacker to potentially capture the transmitted traffic. A user with security awareness and knowledge of the Https protocol which incorporates data encryption, would prefer it for transmitting his credentials.

Twenty-sixth Question This question checks if the user would enter his credit card information on a website which uses the insecure Http protocol. The question is:

Would you enter your credit card information on a web site whose address starts with "http://"?

(e.g. http://www.example.com/)

The user has to choose one of the following options:

- Yes
- No

This question contributes to the user properties of *security oriented education* and *security awareness*. A part of the population perceives only their credit card information as something with monetary value and not their password; this question is expected to assist in the examination of what people might perceive as critical information.

Twenty-seventh Question This question evaluates the user vulnerability by giving a real life example of a social engineering attack. The question is:

You receive by postal mail at your work address a program which is labeled as "critical security updates" and installation instructions. Would you install it?

The user has to choose one of the following options:

- Yes
- No

This question contributes to the user properties of *security awareness* and *responsibility*. At the given example an attacker attempts to take advantage of the user's sense of responsibility towards the system security. A user who receives a program which by first look is identified as something critical to the system security, might install it without considering the origin of it. The success rate of such an attack might be even higher if the package looks genuine and when labelled as it was send by the Information Technology department of the targeted institution.

Twenty-eighth Question This question evaluates the user vulnerability by giving a real life example of a social engineering attack. The question is:

Your work computer fails to connect to the Internet; you receive a call from someone who identifies himself as a network technician and requires your username and password in order to repair your connection. Would you give your username and password over the phone?

The user has to choose one of the following options:

- Yes
- No

This question contributes to the user properties of *security awareness* and *responsibility*. At the given example an attacker attempts to take advantage of a user's need which arises. The attacker contacts by phone the user who has a problem and asks for his credentials in order to give a solution. The user might give away his username and password, driven by the need to solve his problem and from the fact that the attacker called him knowing already that the problem exists. Users are usually not aware of the fact that attackers might fabricate problems in order to proceed in such kind of attacks[12]. The success rate of such an attack would be higher if the attacker would make use of a proper terminology or falsely identifies himself by the name of someone who indeed works at the targeted institution[31]. It is worthy noticing that if the attacker will indeed solve the users problem, then there will be possibly a trust relationship and familiarity between the user and the attacker; where the latter could unauthorized keep asking for favours[12].

Twenty-ninth Question This question evaluates the user vulnerability by giving a real life example of a social engineering attack. The question is:

You have placed some time ago an Internet based order; you receive a call from someone who identifies as salesman from that store and asks you for your credit card information in order to dispatch your purchase. Would you give your credit card information over the phone?

The user has to choose one of the following options:

- Yes
- No

This question contributes to the user properties of *security awareness* and *responsibility*. The given example is similar to the one of the previous question, with the exception that the problem is fabricated. The attacker gained knowledge of the user's order placement and attacks his need of having the order dispatched. The false authentication of the attacker and the fact that he knows details that no one else would normally do, along with the need of the user for having his order completed, might lead to a successful retrieval of the credit card information. The users might respond differently in accordance to their monetary value perception, similarly to the twenty-sixth question.

Installation Environment

The user property of installation environment investigates primarily the users' security alertness in relation to the work environment and secondarily the environment itself.

Thirtieth Question This question checks if the user has the same username for several accounts. The question is:

Do you use a common username for many kinds of accounts?

(e.g. personal account, work account)

The user has to choose one of the following options:

- Yes
- No

This question contributes to the user properties of *security awareness*, *installation environment*, *responsibility* and *compliance*. It is used in the evaluation of all hypotheses as it indicates the security awareness of the user. A common username enhances the ability of an attacker to relate a compromised account with any account linked to the same person. Although this itself cannot be perceived as a security breach, the linking of accounts when one of them is compromised, might make the intrusion to the rest an easier task. Furthermore, the linking of several accounts combined with the usage of a common

password, which is examined on the fifteenth question, is a critical security risk. A recent study had shown that forty percent of online banking users had the same username for other accounts[32]; which makes this bad security practice, a quite frequent one.

Thirty-first Question This question checks if the user needs a password to login to his work computer. The question is:

Do you use a password to login to your work computer?

The user has to choose one of the following options:

- Yes
- No

This question contributes to the user properties of *security awareness*, *installation environment*, *responsibility* and *compliance*. It is used in the evaluation of all hypotheses as it indicates the security awareness of the user. A computer system where it is required to enter a password to login, is considered more secure than one where a password is not required.

Thirty-second Question This question examines if the user has knowledge of the policy on password change frequency. The question is:

How often does your work require you to change your password?

The user has to choose one of the following options:

- At least once per 3 months
- At least once per 6 months
- At least once per 12 months
- Never
- I don't know

This question contributes to the user property of *installation environment*. It is not used in the evaluation of any hypothesis but it adds important knowledge regarding the security policy. A password which changes frequently is less likely to be easily recovered.

Thirty-third Question This question checks if the user's workplace computer is inaccessible by anyone else. The question is:

Is the computer you use at work located in a physically secure location?

(i.e. not accessible by anyone else)

The user has to choose one of the following options:

- Yes
- No

This question contributes to the user properties of *security awareness* and *installation environment*. It is used in the evaluation of the fifth hypothesis as it gives an indication of an attacker's ability to have physical access to the targeted computer system. Once a system's physical security is compromised, there is not any mean for protecting the data; even in the case of implemented encryption, a maliciously[12] installed keylogger or rootkit could capture the user's passphrase.

Thirty-fourth Question This question evaluates if the user would allow someone to use his work computer under his supervision. The question is:

Would you allow someone to use your work computer with your supervision?

The user has to choose one of the following options:

- Yes
- No

This question contributes to the user properties of *security awareness*, *installation environment*, *responsibility* and *compliance*. It is used in the evaluation of all hypotheses as it indicates the security awareness of the user. This question examines the same attributes like the twentieth question but focusing on the work environment.

Thirty-fifth Question This question evaluates if the user would allow someone unsupervised to use his work computer. The question is:

Would you allow someone to use your work computer without your supervision?

The user has to choose one of the following options:

- Yes
- No

This question contributes to the user properties of *security awareness*, *installation environment*, *responsibility* and *compliance*. It is used in the evaluation of all hypotheses as it indicates the security awareness of the user. This question examines the same attributes like the twenty-first question but focusing on the work environment.

[12]Sometimes it also occurs that the employer has intentionally installed such software for monitoring the users' actions

Thirty-sixth Question This question examines if the user works remotely and whether he is aware of any prerequisite security requirements. The question is:

If you do work from home, does your work require use of specific technologies for secure remote access?

(e.g. Virtual Private Network, antivirus software)

The user has to choose one of the following options:

- Yes
- No
- I don't know
- I do not work from home

This question contributes to the user property of *installation environment*. It is not used in the evaluation of any hypothesis but it adds important knowledge regarding the security policy. An installation environment which demands specific security criteria for remote access, is more secure than an environment which offers access without any security mechanisms. It has been found that users who work remotely often engage in behaviour which might expose the system to a threat[16, 33].

Thirty-seventh Question This question checks if the user transfers data between work and home through the usage of removable media. The question is:

Do you use removable media for transferring data between work and home?

(e.g. USB flash drive, optical disc, floppy disk)

The user has to choose one of the following options:

- Yes
- No

This question contributes to the user properties of *security awareness*, *installation environment* and *responsibility*. It is used in the evaluation of the first and fourth hypotheses as it gives a view of security awareness. The usage of removable media between work and home, regardless if the contained data are personal or work related, might introduce security risks on either computer systems. This way of infecting is one of the oldest, since the usage of Internet was not as common as it is nowadays, but yet removable media are still source of many security issues. However, a frequently updated antivirus software, an account without administrative rights and a user who has separate removable media for his work and home, should be an efficient protection for this threat.

Another risk which rises from the usage of removable media, especially when those contain work oriented data, is the one of physical loss. The size and portability of such media introduces less control over the factor of physical ownership. In this case, implementing encryption on the data and keeping a separate backup copy, would be enough to both make the data useless upon loss and still have the original data available.

Thirty-eighth Question This question checks if the user is aware whether his work computer has installed antivirus software. The question is:

Do you have antivirus software installed on your work computer?

The user has to choose one of the following options:

- Yes
- No
- I don't know

This question contributes to the user properties of *security awareness* and *installation environment*. It is used in the evaluation of the first, fourth and fifth hypotheses as it gives a view of security awareness and vulnerability of installation environment. The antivirus software is one of the most basic security software that prevents a system infection from malware. It is suggested that a system configured for a basic security protection, should incorporate an installed antivirus[34].

Security Policy

The user property of security policy examines the policy of the installation environment. This could be a part of the installation environment property but it encapsulates environmental attributes which are, or can be, controlled by the policy. In the case of the thirty-third question, policy cannot control the enviroment if for example a laptop computer is operated; therefore, this question is located at the user property of installation environment.

Thirty-ninth Question This question examines the current policy on antivirus software and whether the user is aware of it. The question is:

Are you required to have antivirus software installed on your work computer?

The user has to choose one of the following options:

- Yes
- No
- I don't know

This question contributes to the user property of *security policy*. It is not used in the evaluation of any hypothesis but it adds important knowledge regarding the security policy. Antivirus software is one of the core defences against malware; a policy with basic security on mind would require the use of antivirus software on the work computer systems.

Fortieth Question This question queries if the user is permitted to install his own software or hardware in his work environment. The question is:

Are you allowed to install your own software or hardware in your work environment?

(e.g. mobile phones connected via bluetooth on your work computer, instant messaging applications, any software that was not pre-installed at your work environment)

The user has to choose one of the following options:

- Yes
- No
- I don't know

This question contributes to the user properties of *security awareness* and *security policy*. It is used in the evaluation of the first, fourth and fifth hypotheses as it gives an insight of the users security awareness and vulnerability surface. By allowing the usage of customised software or hardware at the work environment we allow code execution that is unrestricted and security settings which are set by users. An improperly configured bluetooth device could allow an attacker within the bluetooth signal range to get access; similarly, an instant messaging or music playing software which is not updated, or even malicious from source software[13][35], would be a definite point of intrusion. A security oriented policy would have to disable by default the ability for installation of any external software or hardware, and allow the opposite only in case that it is absolutely required.

Forty-first Question This question checks if the user is required to install his own software of hardware in his work environment. The question is:

Are you required to install your own software or hardware in your work environment?

(e.g. email client software, web browser, web browser extensions)

The user has to choose one of the following options:

- Yes

[13] e.g. illegally distributed software or pirate software

- No
- I don't know

This question contributes to the user property of *security policy*. It is not used in the evaluation of any hypothesis but it adds important knowledge regarding the security policy. If a user is required to install software or hardware in his work environment, then the policy should take the appropriate measures for ensuring the system security. Not all users can be responsible for such a process, but the ones who have the technical and appropriate security oriented knowledge.

Chapter 4

Results Evaluation

This chapter consists of the data coding, interpretation and evaluation. The participants' answers were encoded from the original form into an analyzable format. The interpreted data were later used for creating general observations from the results and evaluating the hypotheses. The hypotheses evaluation will generate new knowledge by either confirmation or refutation of each hypothesis.

The participation from Oslo University College and Domi Educational Group was 59%[1] and 27%[2] respectively. The high difference of participation might be caused from the fact that the survey was distributed to Domi Educational Group in English, while Oslo University College had a Norwegian translation of it.

4.1 Data Coding

Data coding is the process where the questions input is converted to a format which allows data analysis. For the process of data coding the R software was used; every question was interpreted in a way that assigns value when certain criteria are satisfied. As most of the questions were only receiving *Yes and No* answers, the interpretation was usually a boolean function of the form $(d\$q1 ==' Yes')$; where d is the data, $q1$ is the question and $'Yes'$ is the criteria upon which the boolean function is checked. Once a boolean function would be true, it would increase the score on the perspective axis of the analyzed hypothesis. The questions which had a range of values to choose from, were coded by using the sum of multiple boolean functions, with each answer getting a score in relation to the other options; the higher importance an answer would have, the larger the contribution would be on the score of the perspective axis. The categorization of other cases[3] in which the data coding was different, is examined later. The general observations were created by using the $sum()$ command on certain boolean functions and the histograms derived by using the $hist()$ command on each hypothesis axes. The functions which

[1] Oslo University College participation was 294 from a target group of 500

[2] Domi Educational Group participation was 40 from a target group of 150

[3] i.e. password scoring system, uncertainty answers

were used for the evaluation of the hypotheses can be found in Table 4.1.

4.1.1 Password Scoring System

The twelfth and thirteenth question examine the user's knowledge on generating a strong password; the categorization of the user's answers was made through creating a password scoring system. Each password characteristic would either add no score, or add a given score when the user's choices satisfy certain criteria. The password characteristics were separated into very good, good and bad. Using phrases for a password generation was chosen as very good password characteristic which when encountered, the user's password score would increase by two. As good password characteristics were chosen the following:

- Not using dictionary words
- Not using personal information
- Combining lowercase and uppercase letters
- Using symbols
- Using numbers
- Be at least 8 or more characters long

When each of the above good characteristics would be encountered the user's password score would increase by one. Any other choice would result on a poor password generation and therefore it would be given no score.

4.1.2 Uncertainty Answers

The reason for adding answer options[4] which include uncertainty, was because people could be unaware or not sure about something. However, uncertainty in Information Security can be seen as possibility of a system exposure[5]. As near misses are like accidents which are waiting to happen, any answer which incorporated uncertainty had the score of the most related answer with certainty. For example, any answer of a user on sixteenth question except the negative one would mean that there is a possibility of writing down a password that is difficult to remember.

4.2 Data Interpretation

During the process of data interpretation the quantified data are used for the hypotheses evaluation. The first and second dimension of each hypothesis

[4] i.e. I don't know, maybe

q1	N/A
q2	N/A
q3	(d$q3=='Less than 3 hours') + (d$q3=='3-8 hours')*2 + (d$q3=='More than 8 hours')*3
q4	(d$q4=='Less than once per week') + (d$q4=='At least once per week')*2 + (d$q4=='At least once per day')*3 + (d$q4=='At least once per hour')*4
q5	(d$q5=='Yes')
q6	(d$q6=='Yes')
q7	(d$q7=='5') + (d$q7=='4')*2 + (d$q7=='3')*3 + (d$q7=='2')*4 + (d$q7=='1')*5
q8	(d$q8=='5') + (d$q8=='4')*2 + (d$q8=='3')*3 + (d$q8=='2')*4 + (d$q8=='1')*5
q9	(d$q9=='Yes')
q10	(d$q10=='Yes')
q11	(d$q11=='Yes')
q12	(grepl('Phrases',d$q12))*2 + (!(grepl('Dictionary words',d$q12))) + (!(grepl('Personal Information',d$q12))) + ((grepl('Lowercase letters',d$q12)) && (grepl('Uppercase letters',d$q12))) + (grepl('Symbols',d$q12)) + (grepl('Numbers',d$q12))
q13	(d$q13=='8 or more characters')
q14	(d$q14=='Never') + (d$q14=='Only when required by the system')*2 + (d$q14=='At least once per 12 months')*3 + (d$q14=='At least once per 6 months')*4 + (d$q14=='At least once per 3 months')*5
q15	(d$q15=='No')
q16	(d$q16=='No')
q17	(d$q17=='No')
q18	(d$q18=='Yes')
q19	(d$q19=='Yes')
q20	(d$q20=='No')
q21	(d$q21=='No')
q22	(d$q22=='No')
q23	(d$q23=='No')
q24	(d$q24=='No')
q25	(d$q25=='No')
q26	(d$q26=='No')
q27	(d$q27=='No')
q28	(d$q28=='No')
q29	(d$q29=='No')
q30	(d$q30=='No')
q31	(d$q31=='Yes')
q32	N/A
q33	(d$q33=='Yes')
q34	(d$q34=='No')
q35	(d$q35=='No')
q36	N/A
q37	(d$q37=='No')
q38	(d$q38=='Yes')
q39	N/A
q40	(d$q40=='No')
q41	N/A

Table 4.1: Functions for the Hypotheses Evaluation

would serve as independent and dependent variables respectively; both variables would receive score from the particular to them questions[5]. The variables are plotted in a two-dimensional diagram which visualise the score distribution. In the plots, each point represents a user's score; the more scores a point would sum, the larger it would appear. For enhancing the data interpretation, the axes histograms are plotted as well. The purpose of plotting is to find trends on data distribution which would be useful on the hypotheses evaluation. In addition, the median score was used for measuring the central tendency of the axes; which was found particularly useful in comparing the responses from the two participating institutions.

4.3 General Observations

Apart from the hypotheses evaluation, the survey can offer a highly informative input when the questions are seen either individually or in relation to each other. As a comprehensive evaluation of all possible questions associations would be a highly time consuming process, the purpose of this section is to view a noteworthy part of these results. The developed R code which was used for this section can be found in the Appendices. Table 4.2 lists vertically the cases and horizontally the results; percentages were calculated in relation to the total participation and results got rounded to the nearest integer.

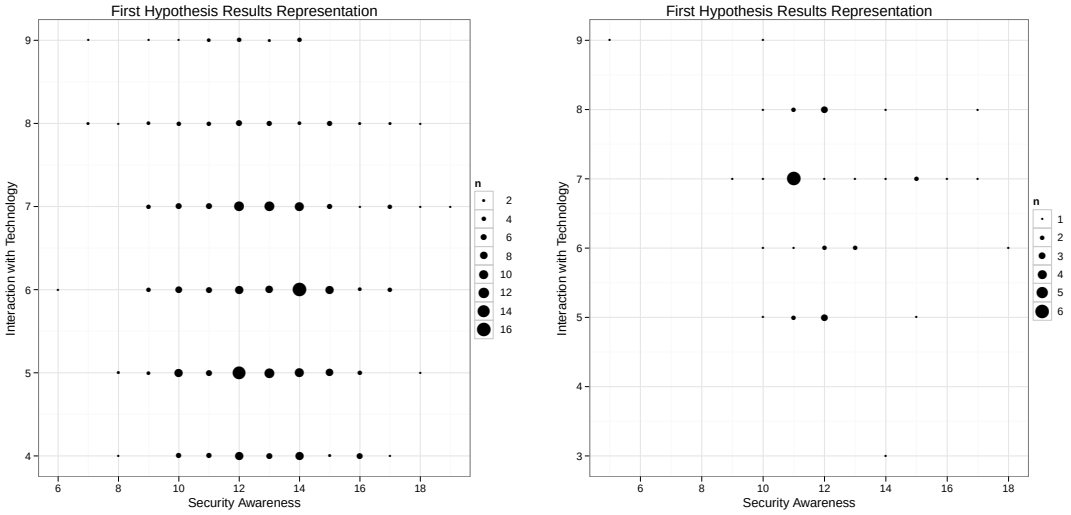
[5]See Figure 3.2

	Oslo University College		Domi Educational Group	
	Number of Answers	Percentage	Number of Answers	Percentage
Users who had never received security training	216	73%	23	58%
Users who generate on average passwords which are less than eight characters	135	46%	16	40%
Users who use a common username and password for several accounts	125	43%	25	63%
Users who would write down a complicated password	261	89%	28	70%
Users who would open an email link or attachment from an address they recognize	267	91%	37	93%
Users who would open an email link or attachment from an address they do not recognize	38	13%	7	18%
Users who would share their credentials with someone else	19	6%	13	33%
Users who would enter their credentials or credit card information on an insecure website	80	27%	22	55%
Users who would enter their credentials and credit card information on an insecure website	17	6%	8	20%
Users who would allow someone to use their work computer with their supervision	269	91%	30	75%
Users who would allow someone to use their work computer without their supervision	121	41%	12	30%
Users who would allow someone to use their work computer with their supervision, but not their home computer	12	4%	2	5%
Users who would allow someone to use their work computer without their supervision, but not their home computer	21	7%	1	3%
Users who transfer data between work and home, and do not have an antivirus installed on their home computer	11	4%	3	8%
Users who are required to install software or hardware, and do not have professional skills	24	8%	12	30%
Users who are allowed to install their own software or hardware at work	77	26%	14	35%

Table 4.2: General Observations

4.4 First Hypothesis

The first hypothesis examines whether people with higher technology interaction show increased security awareness. From the given results in Figure 4.1a a diagonal distribution can be seen which can be also matched with the equivalent histogram in Figure 4.2, where the values are lowered in the centre of the axis. For Oslo University College the overall distribution of values shows less interaction with technology and higher security awareness in comparison with the one from Domi Educational Group. However, the limited responses of Domi Educational Group would not allow to derive any conclusions.



(a) Oslo University College

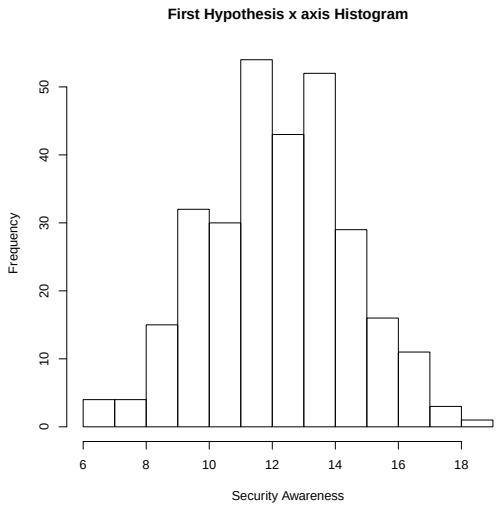
(b) Domi Educational Group

Figure 4.1: First Hypothesis Results Representation

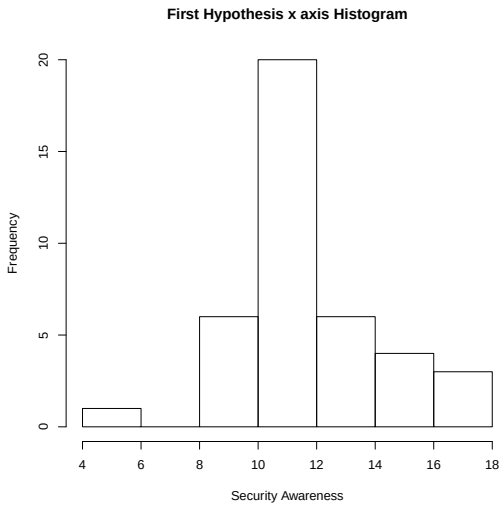
According to Table 4.3, the median values of both institutions results seem to be close; comparing the histograms of both axes seem to show a difference which most probably occur due to the limited sample of Domi Educational Group. The particularly high difference can be spotted by closely examining Figure 4.3b, where the distributions are different with most of the results being on the rightmost side; having the leftmost side rather limited in comparison with Figure 4.3a. As there is evidence of interaction with technology increasing in parallel with security awareness, the hypothesis is confirmed.

Oslo University College		Domi Educational Group	
x axis	y axis	x axis	y axis
13	6	12	7

Table 4.3: Median Values for the axes of First Hypothesis

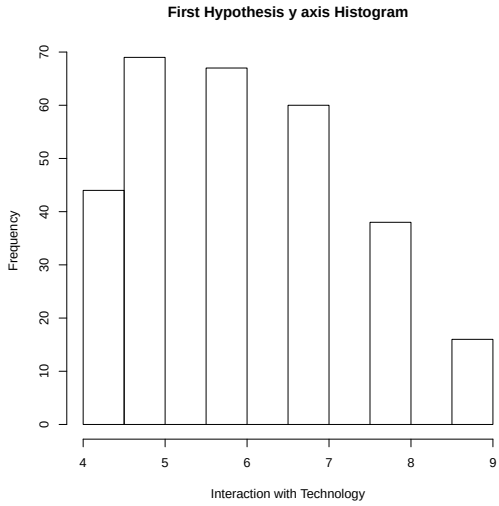


(a) Oslo University College

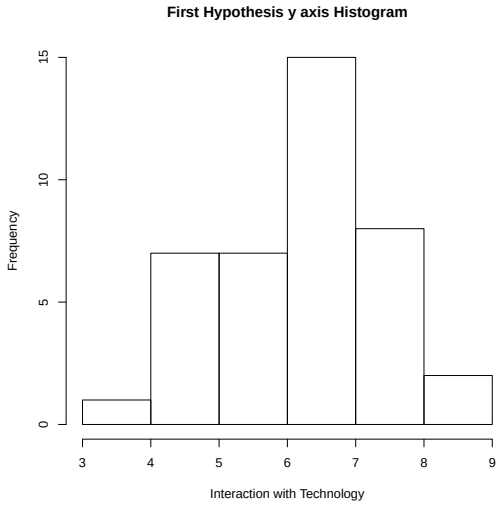


(b) Domi Educational Group

Figure 4.2: First Hypothesis x axis Histogram



(a) Oslo University College

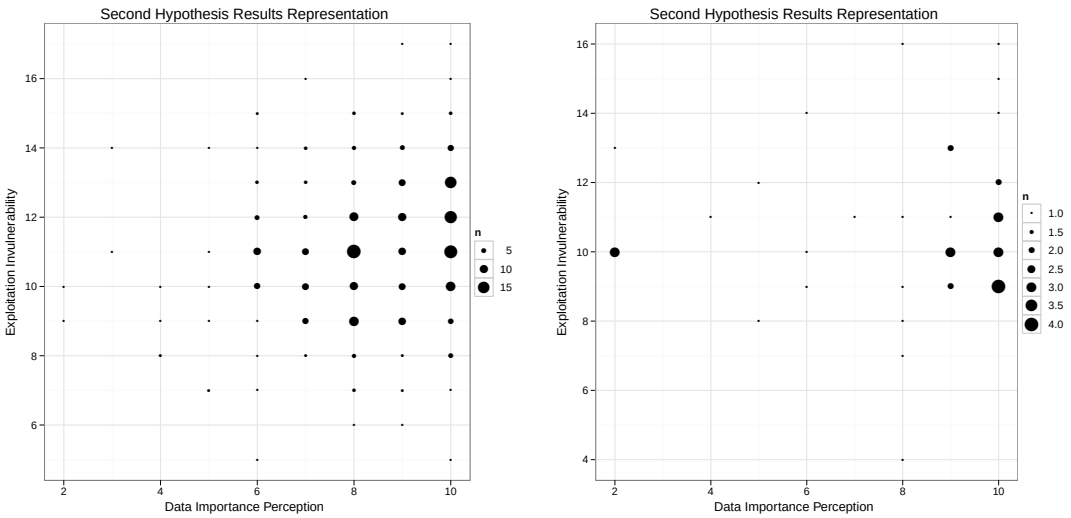


(b) Domi Educational Group

Figure 4.3: First Hypothesis y axis Histogram

4.5 Second Hypothesis

The second hypothesis evaluates if the population who perceives personal and work data as more important to protect, is less vulnerable to attacks. The derived results in Figure 4.4a show that the rightmost part with value 10, which is the part where people consider both work and home data as extremely important, is on average less vulnerable than the rest of the answers; something which does not appear to exist at the results from Domi Educational Group as seen in Figure 4.4b.



(a) Oslo University College

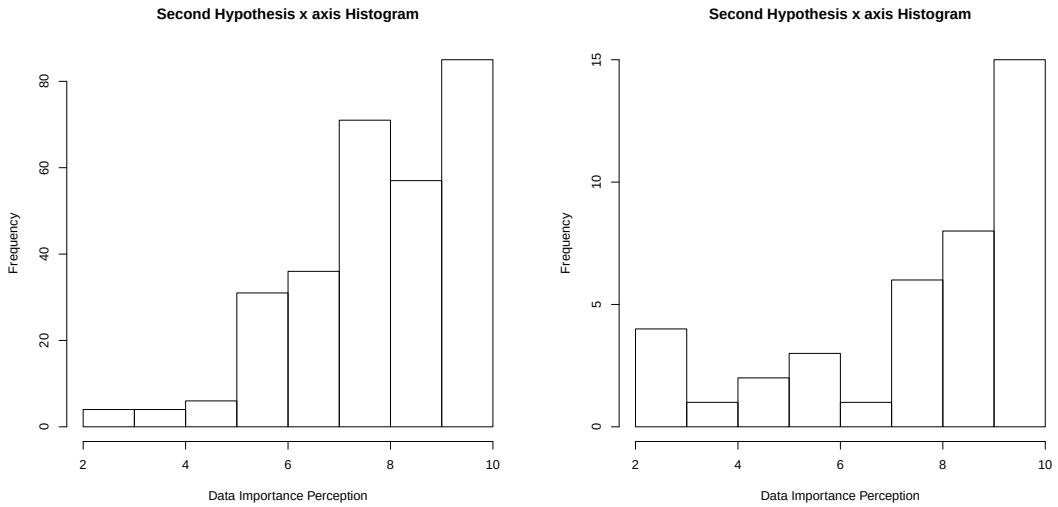
(b) Domi Educational Group

Figure 4.4: Second Hypothesis Results Representation

The data from Domi Educational Group show some irregularities when compared to the ones from Oslo University College. Although the axes median values do not significantly differ as the Table 4.4 indicates, according to Figures 4.5 and 4.6 the distribution of scores is quite irregular in comparison with Oslo University College. A possible reason for this might be the smaller participation from Domi Educational Group. As the results in Figure 4.4a show, the hypothesis is confirmed because people with higher perception of data importance show a higher invulnerability level than all the rest.

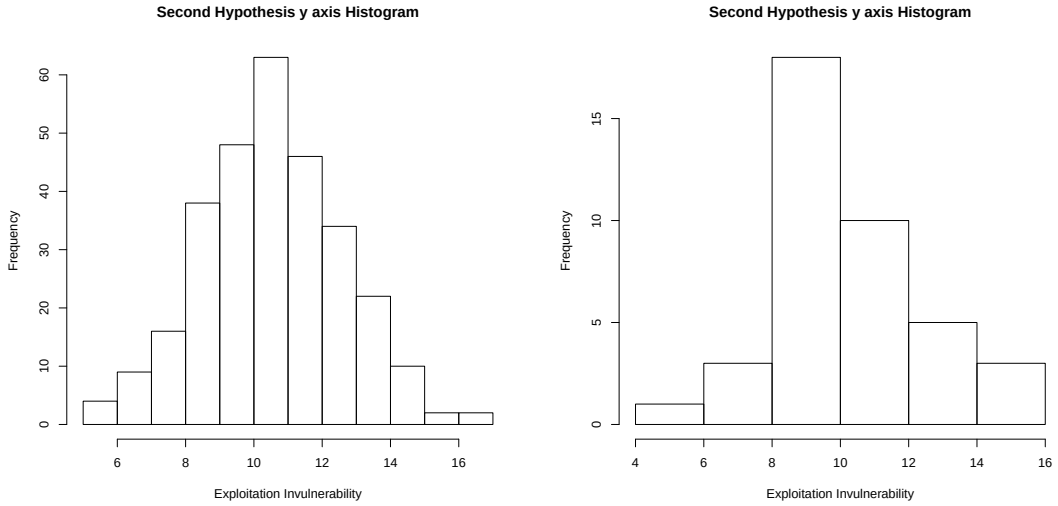
Oslo University College		Domi Educational Group	
x axis	y axis	x axis	y axis
8	11	9	10

Table 4.4: Median Values for the axes of Second Hypothesis



(a) Oslo University College (b) Domi Educational Group

Figure 4.5: Second Hypothesis x axis Histogram

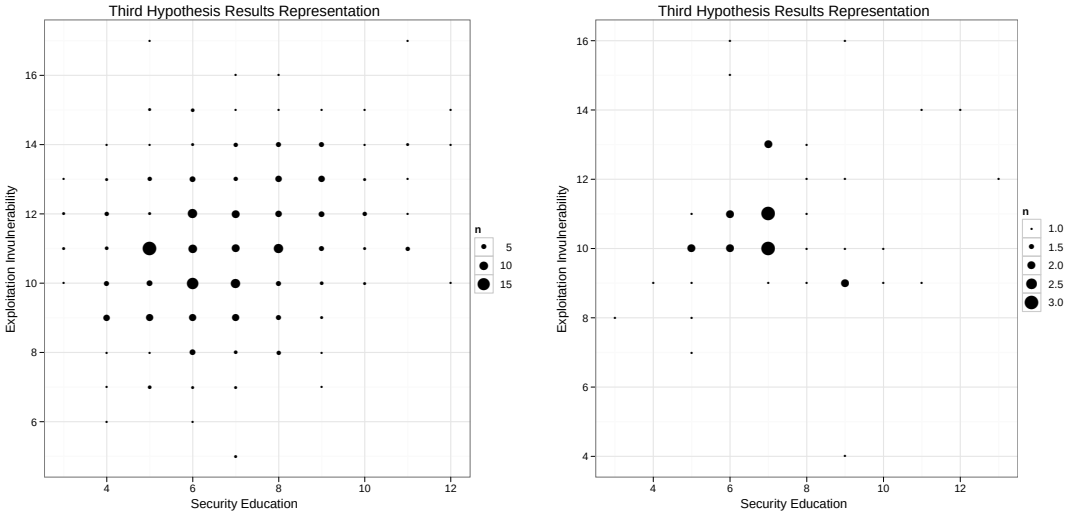


(a) Oslo University College (b) Domi Educational Group

Figure 4.6: Second Hypothesis y axis Histogram

4.6 Third Hypothesis

The third hypothesis examines if population with a higher security education would be less vulnerable to attacks. By having a look at Figure 4.7a one can see that the results appear to be spread diagonally. Both institutions appear to have a similar result distribution but the limited resolution of the results from Domi Educational Group would again prevent deriving any conclusions.



(a) Oslo University College

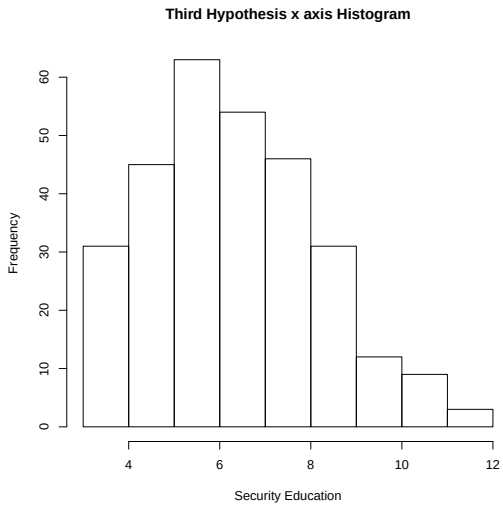
(b) Domi Educational Group

Figure 4.7: Third Hypothesis Results Representation

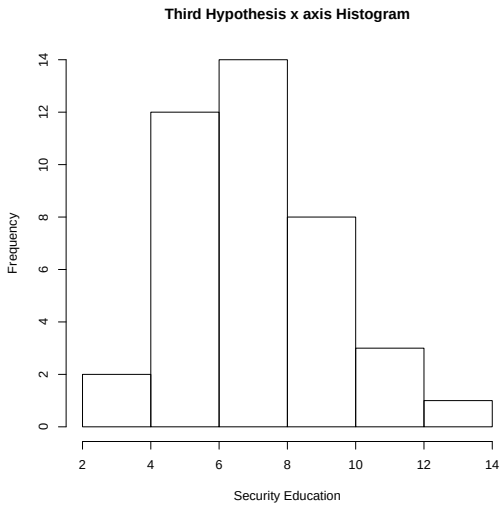
Both institutions median values are very close as seen from Table 4.5. By comparing the histograms in Figure 4.8 one can see that they have an almost identical pattern which differs only at the resolution. The noticeable difference between the two institutions results surfaces when comparing the histogram found in Figure 4.9; while Figure 4.9a shows an approximately normal curve distribution, Figure 4.9b seems to have less results on the left part of the histogram; something which also excuses the smaller median value for the y axis of Domi Educational Group found at Table 4.5. As the Figure 4.7a shows a diagonal spread of the results, it is confirmed that to some extent, higher security education correlates with higher exploitation invulnerability.

Oslo University College		Domi Educational Group	
x axis	y axis	x axis	y axis
7	11	7	10

Table 4.5: Median Values for the axes of Third Hypothesis

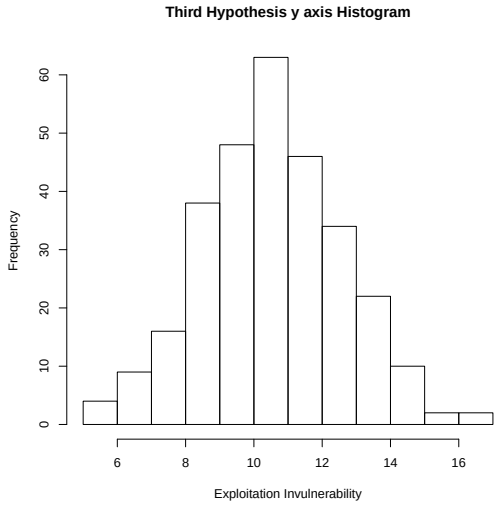


(a) Oslo University College

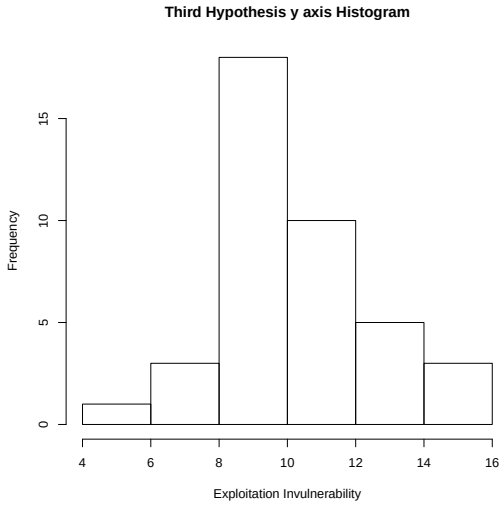


(b) Domi Educational Group

Figure 4.8: Third Hypothesis x axis Histogram



(a) Oslo University College

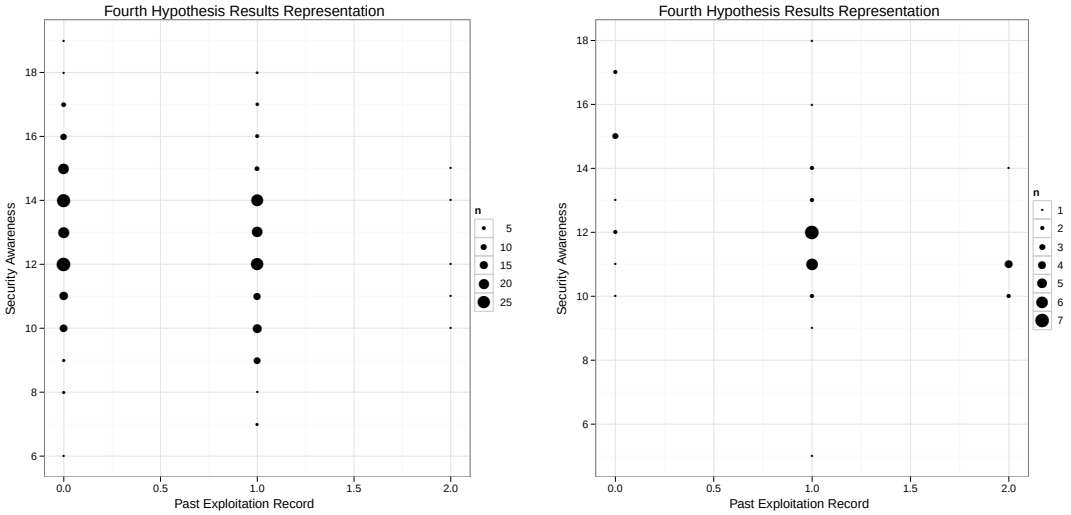


(b) Domi Educational Group

Figure 4.9: Third Hypothesis y axis Histogram

4.7 Fourth Hypothesis

The fourth hypothesis examines whether people who experienced a security incident in the past are more security-aware. Figure 4.10a demonstrates the opposite result than the original assumption; people who had no prior security incidents seem to be on average more security-aware than people who had one or both types of the examined security incidents. A similar trend is demonstrated from the results of Domi Educational Group in Figure 4.10b, although with a reduced resolution due to limited responses.



(a) Oslo University College

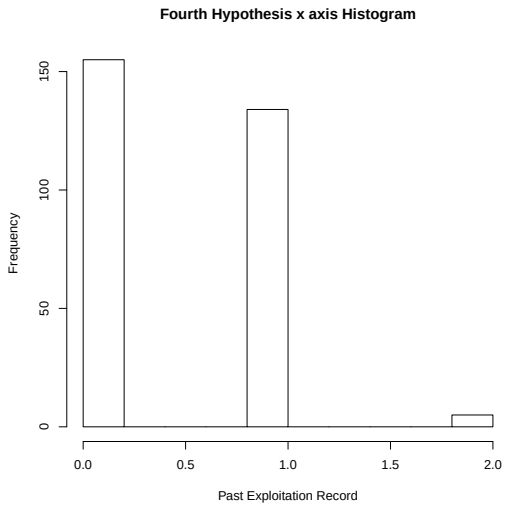
(b) Domi Educational Group

Figure 4.10: Fourth Hypothesis Results Representation

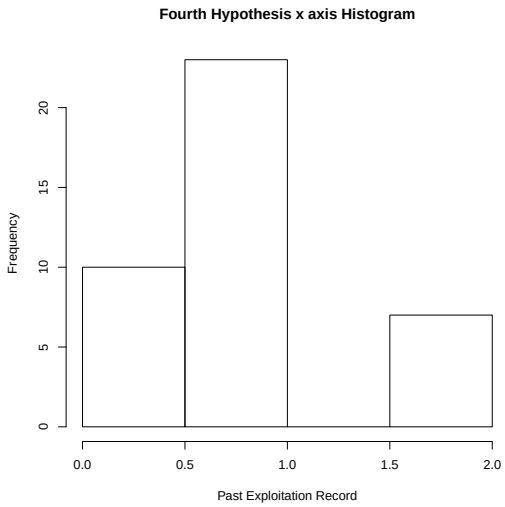
Although the median values of both institutions, as seen in Table 4.6, do not appear to have a large difference, the distribution of both axes values differ significantly. By comparing the Figures 4.11a and 4.11b one can see that the results of the latter for value 0 are rather limited proportionally to the first; something which also gets visually confirmed by Figure 4.10b. The histograms in Figure 4.12 could have a similar distribution, but due to the limited resolution of Figure 4.12b it is not possible to derive any conclusion. As the results of Oslo University College in Figure 4.10a show lower security awareness for people with prior security incidents, the hypothesis is not supported.

Oslo University College		Domi Educational Group	
x axis	y axis	x axis	y axis
0	13	1	12

Table 4.6: Median Values for the axes of Fourth Hypothesis

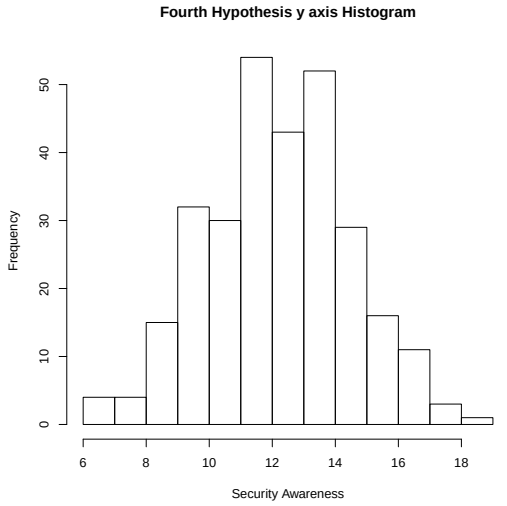


(a) Oslo University College

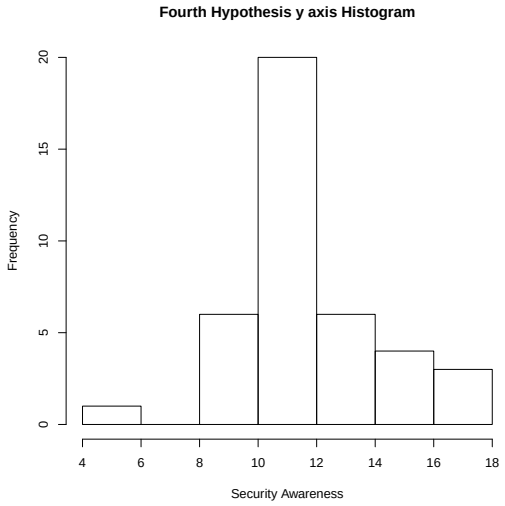


(b) Domi Educational Group

Figure 4.11: Fourth Hypothesis x axis Histogram



(a) Oslo University College

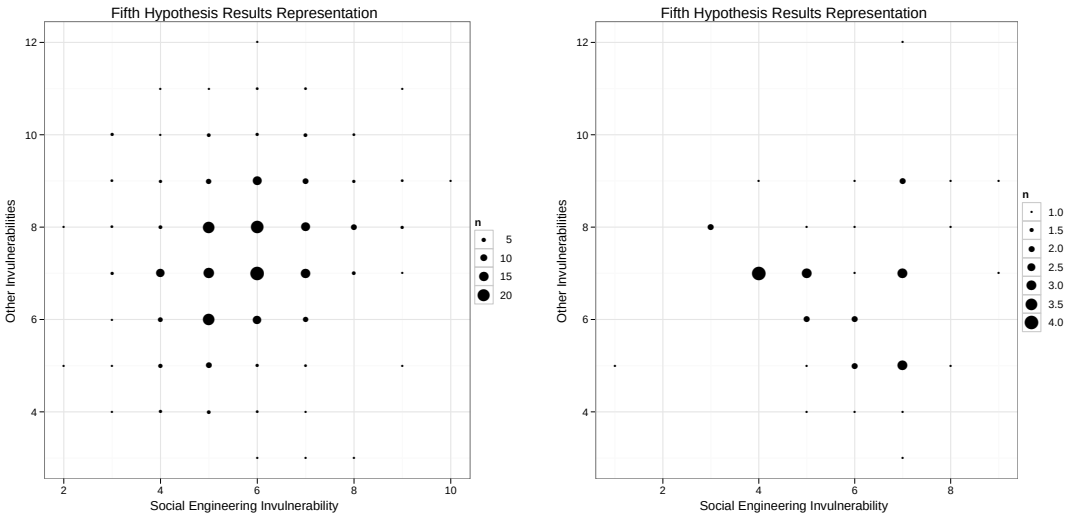


(b) Domi Educational Group

Figure 4.12: Fourth Hypothesis y axis Histogram

4.8 Fifth Hypothesis

The fifth hypothesis evaluates whether invulnerability to social engineering does not correlate with other invulnerabilities. An initial look in Figures 4.13a and 4.13b shows an opposite diagonal orientation for each. While in Figure 4.13a the other invulnerabilities show a relative increase in relation to higher values of social engineering invulnerability, Figure 4.13b shows a decrease.



(a) Oslo University College

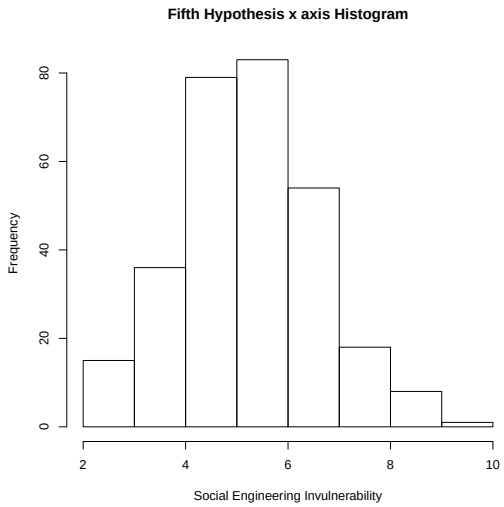
(b) Domi Educational Group

Figure 4.13: Fifth Hypothesis Results Representation

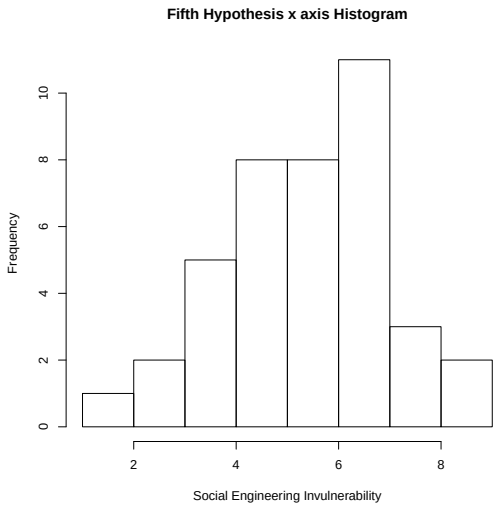
The median values in Table 4.7 show the same results for both institutions; however, the distribution of values as seen in Figures 4.14 and 4.15 looks different between the two participating institutions. Figure 4.15b has a particularly unusual pattern in relation to Figure 4.15a. As the results do not demonstrate evidence of correlation between social engineering and other invulnerabilities, the hypothesis is confirmed.

Oslo University College		Domi Educational Group	
x axis	y axis	x axis	y axis
6	7	6	7

Table 4.7: Median Values for the axes of Fifth Hypothesis

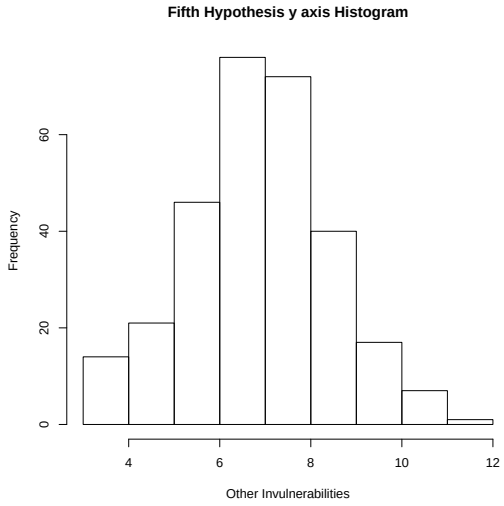


(a) Oslo University College

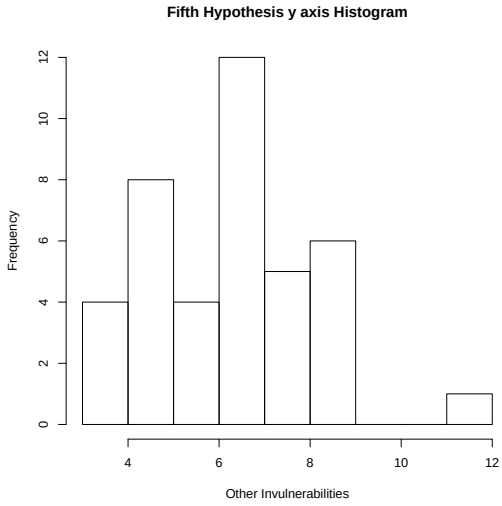


(b) Domi Educational Group

Figure 4.14: Fifth Hypothesis x axis Histogram



(a) Oslo University College



(b) Domi Educational Group

Figure 4.15: Fifth Hypothesis y axis Histogram

Chapter 5

Future Suggestions

This research answer a question which could attract further study or conduct an evaluation from other perspectives; the offered knowledge can be the basis for further investigation of the human factor in another setting. A noteworthy enhancement on the current study would be the addition of weighting factors to each question; a possibility which was not implemented due to time constraints, but would result on a more concrete output. As not all questions have the same importance or possibility to occur, it would be more consistent to evaluate them in regards to each other as well, and not only to add their score in a co-variable. Additionally, it would be of a high interest, but ethically questionable, to investigate the human factor vulnerability by an on-site exploitation in a controlled environment[1]; this method would significantly reduce the error probability and would give a better insight of the human factor status. Finally, an alternative approach by applying a qualitative or mixed research method, could give surface to the reasons behind peoples' actions.

Some future considerations which adjust the followed topic, would be to additionally discuss the intentional attack of one on his own premises; a study which would result to an overall evaluation of the human factor, regardless of the attack source and nature. The human factor in Information Security could be also examined in relation to the dimensions of productivity, cost and usability; dimensions particularly interesting for a corporate environment.

[1] Always with the granted permission of the respective administration

Chapter 6

Conclusion

Amateurs hack systems, professionals hack people.
Bruce Schneier

Human factors can have a high impact in Information Security; the current study investigated the influence of human factors in regards to accidental exposure. In confirmation of past research, findings prove that a high number of users would engage into actions which could make the system subject to an attack. Users' behaviour have been shown to be related to technology interaction, data importance perception and security oriented education; while the presence of a previous security incident, does not indicate that people would be more aware of security threats in the future. Furthermore, there is evidence that users who are invulnerable to various types of attacks, are not necessarily also invulnerable to social engineering attacks; something initially expected as the latter are harder to detect and respond.

The answer on the human factor security issues can be supplied by education. Security oriented education can address the human factor problem by increasing the users' practical[1] and theoretical[2] knowledge. Similarly, the data importance perception could be used in addressing the human factor by increasing the users' feeling of responsibility; users would then understand why security is necessary in the given circumstances. Additionally, attention should be given on addressing the social engineering vulnerabilities; such attacks exploit directly the human factor with a high effectiveness. Users would have to receive a series of social engineering case studies as examples, and training on how to authenticate people.

As long as Information Technology systems have the human as a prerequisite component, a responsible security evaluation process should without doubt include users. With both technology and people evolving by time, security should parallely adopt and progress according to the new advances. Even under the ideal thought of a faultless system, people would still be subject to an error and therefore, a potential point of intrusion.

[1] e.g. password generation
[2] e.g. recognizing attacks

Appendices

Appendix G

Informed Consent

This is a survey which evaluates the human factors in information security; it is part of a research study made by Theodoros Nikolakopoulos, master student at Oslo University College and University of Oslo. It has been approved by [1]. The distribution of the survey at the institution staff has been made by the IT Department.
The Terms and Conditions of the survey are the following:

1. Non-coercion
By completing the following questionnaire and pressing the "Send" button you confirm that you willingly participate in the research study.

2. Confidentiality
The gathered data will be kept strictly confidential and the only one who will have access will be the one who conducts the research study. The participants and the institution may have access only on the publicly available research study results upon request, and not on the individual data.

3. Maintenance of Privacy
The participation in this survey remain by all means anonymous. By the submitted answers it will not be possible to identify any individual.

4. Protection From Harm
The current research study will not have any physical harm on the participants. If you do believe that you might suffer a psychological harm from the completion of this survey, please avoid from doing so and feel free to submit your feedback by using one of the below mentioned emails.

For any related questions please contact the research study conductor at:

Theodoros.Nikolakopoulos@stud.iu.hio.no
TheodoN@ifi.uio.no

[1] Konstantinos Katsifis, for Domi Educational Group; Åsulv Frøysnes and Ole Lycke, for Oslo University College

Appendix H

R code for General Observations

```
d <- read.csv('data_no.csv')
#Read the survey data

sum(d$q11=='No')
#Users who had never received security training

sum(d$q13!='8 or more characters')
#Users who generate on average passwords which are less
    than eight characters

sum(d$q15[d$q30=='Yes']=='Yes')
#Users who use a common username and password for several
     accounts

sum(d$q16!='No')
#Users who would write down a complicated password

sum(d$q22=='Yes')
#Users who would open an email link or attachment from an
     address they recognize

sum(d$q23=='Yes')
#Users who would open an email link or attachment from an
     address they do not recognize

sum(d$q24=='Yes')
#Users who would share their credentials with someone
    else

sum(((d$q25=='Yes')+(d$q26=='Yes'))!='0')
#Users who would enter their credentials or credit card
    information on an insecure website

```

```
sum(((d$q25=='Yes')+(d$q26=='Yes'))=='2')
#Users who would enter their credentials and credit card
    information on an insecure website

sum(d$q34=='Yes')
#Users who would allow someone to use their work computer
     with their supervision

sum(d$q35=='Yes')
#Users who would allow someone to use their work computer
     without their supervision

sum(d$q34[d$q20=='No']=='Yes')
#Users who would allow someone to use their work computer
     with their supervision, but not their home computer

sum(d$q35[d$q21=='No']=='Yes')
#Users who would allow someone to use their work computer
     without their supervision, but not their home
    computer

sum(d$q37[d$q19!='Yes']=='Yes')
#Users who transfer data between work and home, and do
    not have an antivirus installed on their home computer

sum(d$q41[d$q1!='Professional␣(e.g.␣administering␣systems
    ␣for␣other␣users)']=='Yes')
#Users who are required to install software or hardware,
    and do not have professional skills

sum(d$q40=='Yes')
#Users who are allowed to install their own software or
    hardware at work
```

Appendix I

R code for Hypotheses Evaluation

```
1  library(ggplot2)
2  #Load the ggplot2 library
3
4  theme_set(theme_bw())
5  #Set the theme to Black/White for ggplot2
6
7  d <- read.csv('data_no.csv')
8  #Read the survey data
9
10 hyp1x_no <- (d$q15=='No') + (d$q16=='No') + (d$q17=='No')
       + (d$q18=='Yes') + (d$q19=='Yes') + (d$q20=='No') + (
     d$q21=='No') + (d$q22=='No') + (d$q23=='No') + (d$q24
     =='No') + (d$q25=='No') + (d$q26=='No') + (d$q27=='No'
     ) + (d$q28=='No') + (d$q29=='No') + (d$q30=='No') + (d
     $q31=='Yes') + (d$q34=='No') + (d$q35=='No') + (d$q37
     =='No') + (d$q38=='Yes') + (d$q40=='No')
11 #Calculate x for the first hypothesis
12
13 hyp1y_no <- (d$q3=='Less_than_3_hours') + (d$q3=='3-8_
     hours')*2 + (d$q3=='More_than_8_hours')*3 + (d$q4=='
     Less_than_once_per_week') + (d$q4=='At_least_once_per_
     week')*2 + (d$q4=='At_least_once_per_day')*3 + (d$q4==
     'At_least_once_per_hour')*4 + (d$q5=='Yes') + (d$q6=='
     Yes')
14 #Calculate y for the first hypothesis
15
16 hyp2x_no <- (d$q7=='5') + (d$q7=='4')*2 + (d$q7=='3')*3 +
       (d$q7=='2')*4 + (d$q7=='1')*5 + (d$q8=='5') + (d$q8==
     '4')*2 + (d$q8=='3')*3 + (d$q8=='2')*4 + (d$q8=='1')*5
17 #Calculate x for the second hypothesis
18
```

```
hyp2y_no <- (d$q15=='No') + (d$q16=='No') + (d$q17=='No') + (d$q18=='Yes') + (d$q19=='Yes') + (d$q20=='No') + (d$q21=='No') + (d$q22=='No') + (d$q23=='No') + (d$q24=='No') + (d$q25=='No') + (d$q26=='No') + (d$q27=='No') + (d$q28=='No') + (d$q29=='No') + (d$q30=='No') + (d$q31=='Yes') + (d$q34=='No') + (d$q35=='No')
#Calculate y for the second hypothesis

hyp3x_no <- (d$q11=='Yes') + (grepl('Phrases',d$q12))*2 + (!(grepl('Dictionary words',d$q12))) + (!(grepl('Personal Information',d$q12))) + ((grepl('Lowercase letters',d$q12)) && (grepl('Uppercase letters',d$q12))) + (grepl('Symbols',d$q12)) + (grepl('Numbers',d$q12)) + (d$q13=='8 or more characters') + (d$q14=='Never') + (d$q14=='Only when required by the system')*2 + (d$q14=='At least once per 12 months')*3 + (d$q14=='At least once per 6 months')*4 + (d$q14=='At least once per 3 months')*5
#Calculate x for the third hypothesis

hyp3y_no <- (d$q15=='No') + (d$q16=='No') + (d$q17=='No') + (d$q18=='Yes') + (d$q19=='Yes') + (d$q20=='No') + (d$q21=='No') + (d$q22=='No') + (d$q23=='No') + (d$q24=='No') + (d$q25=='No') + (d$q26=='No') + (d$q27=='No') + (d$q28=='No') + (d$q29=='No') + (d$q30=='No') + (d$q31=='Yes') + (d$q34=='No') + (d$q35=='No')
#Calculate y for the third hypothesis

hyp4x_no <- (d$q9=='Yes') + (d$q10=='Yes')
#Calculate x for the fourth hypothesis

hyp4y_no <- (d$q15=='No') + (d$q16=='No') + (d$q17=='No') + (d$q18=='Yes') + (d$q19=='Yes') + (d$q20=='No') + (d$q21=='No') + (d$q22=='No') + (d$q23=='No') + (d$q24=='No') + (d$q25=='No') + (d$q26=='No') + (d$q27=='No') + (d$q28=='No') + (d$q29=='No') + (d$q30=='No') + (d$q31=='Yes') + (d$q34=='No') + (d$q35=='No') + (d$q37=='No') + (d$q38=='Yes') + (d$q40=='No')
#Calculate y for the fourth hypothesis

hyp5x_no <- (d$q20=='No') + (d$q21=='No') + (d$q22=='No') + (d$q23=='No') + (d$q24=='No') + (d$q27=='No') + (d$q28=='No') + (d$q29=='No') + (d$q34=='No') + (d$q35=='No')
#Calculate x for the fifth hypothesis

hyp5y_no <- (d$q15=='No') + (d$q16=='No') + (d$q17=='No')
```

```
    + (d$q18=='Yes') + (d$q19=='Yes') + (d$q25=='No') + (
    d$q26=='No') + (d$q30=='No') + (d$q31=='Yes') + (d$q33
    =='Yes') + (d$q38=='Yes') + (d$q40=='No')
#Calculate y for the fifth hypothesis

ggplot(data.frame(x=hyp1x_no, y=hyp1y_no), aes(x = x, y =
    y)) + stat_sum(aes(size = ..n..)) + opts(title='First
    _Hypothesis_Results_Representation') + xlab('Security_
    Awareness') + ylab('Interaction_with_Technology')
ggsave(filename='hyp1_no.pdf', dpi=600)
#Output to pdf the first hypothesis results

ggplot(data.frame(x=hyp2x_no, y=hyp2y_no), aes(x = x, y =
    y)) + stat_sum(aes(size = ..n..)) + opts(title='
    Second_Hypothesis_Results_Representation') + xlab('
    Data_Importance_Perception') + ylab('Exploitation_
    Invulnerability')
ggsave(filename='hyp2_no.pdf', dpi=600)
#Output to pdf the second hypothesis results

ggplot(data.frame(x=hyp3x_no, y=hyp3y_no), aes(x = x, y =
    y)) + stat_sum(aes(size = ..n..)) + opts(title='Third
    _Hypothesis_Results_Representation') + xlab('Security_
    Education') + ylab('Exploitation_Invulnerability')
ggsave(filename='hyp3_no.pdf', dpi=600)
#Output to pdf the third hypothesis results

ggplot(data.frame(x=hyp4x_no, y=hyp4y_no), aes(x = x, y =
    y)) + stat_sum(aes(size = ..n..)) + opts(title='
    Fourth_Hypothesis_Results_Representation') + xlab('
    Past_Exploitation_Record') + ylab('Security_Awareness'
    )
ggsave(filename='hyp4_no.pdf', dpi=600)
#Output to pdf the fourth hypothesis results

ggplot(data.frame(x=hyp5x_no, y=hyp5y_no), aes(x = x, y =
    y)) + stat_sum(aes(size = ..n..)) + opts(title='Fifth
    _Hypothesis_Results_Representation') + xlab('Social_
    Engineering_Invulnerability') + ylab('Other_
    Invulnerabilities')
ggsave(filename='hyp5_no.pdf', dpi=600)
#Output to pdf the fifth hypothesis results

pdf(file='hyp1x_no.pdf')
hist(hyp1x_no, main='First_Hypothesis_x_axis_Histogram',
    xlab='Security_Awareness')
dev.off()
```

```
63 #Output to pdf the first hypothesis x axis histogram
64
65 pdf(file='hyp1y_no.pdf')
66 hist(hyp1y_no, main='First Hypothesis y axis Histogram',
     xlab='Interaction with Technology')
67 dev.off()
68 #Output to pdf the first hypothesis y axis histogram
69
70 pdf(file='hyp2x_no.pdf')
71 hist(hyp2x_no, main='Second Hypothesis x axis Histogram',
      xlab='Data Importance Perception')
72 dev.off()
73 #Output to pdf the second hypothesis x axis histogram
74
75 pdf(file='hyp2y_no.pdf')
76 hist(hyp2y_no, main='Second Hypothesis y axis Histogram',
      xlab='Exploitation Invulnerability')
77 dev.off()
78 #Output to pdf the second hypothesis y axis histogram
79
80 pdf(file='hyp3x_no.pdf')
81 hist(hyp3x_no, main='Third Hypothesis x axis Histogram',
     xlab='Security Education')
82 dev.off()
83 #Output to pdf the third hypothesis x axis histogram
84
85 pdf(file='hyp3y_no.pdf')
86 hist(hyp3y_no, main='Third Hypothesis y axis Histogram',
     xlab='Exploitation Invulnerability')
87 dev.off()
88 #Output to pdf the third hypothesis y axis histogram
89
90 pdf(file='hyp4x_no.pdf')
91 hist(hyp4x_no, main='Fourth Hypothesis x axis Histogram',
      xlab='Past Exploitation Record')
92 dev.off()
93 #Output to pdf the fourth hypothesis x axis histogram
94
95 pdf(file='hyp4y_no.pdf')
96 hist(hyp4y_no, main='Fourth Hypothesis y axis Histogram',
      xlab='Security Awareness')
97 dev.off()
98 #Output to pdf the fourth hypothesis y axis histogram
99
100 pdf(file='hyp5x_no.pdf')
101 hist(hyp5x_no, main='Fifth Hypothesis x axis Histogram',
     xlab='Social Engineering Invulnerability')
```

```
dev.off()
#Output to pdf the fifth hypothesis x axis histogram

pdf(file='hyp5y_no.pdf')
hist(hyp5y_no, main='Fifth Hypothesis y axis Histogram',
    xlab='Other Invulnerabilities')
dev.off()
#Output to pdf the fifth hypothesis y axis histogram
```

Bibliography

[1] Shari Lawrence Pfleeger Charles P. Pfleeger. *Security in Computing*. Pearson Education, Inc., third edition, 2003.

[2] Matt Bishop. *Computer Security: Art and Science*. Pearson Education, Inc., 2003.

[3] D. Ian Hopper. Hackers target university databases, June 2001.

[4] Andrew Brandt Kim Zetter. How hackers hack, April 2001.

[5] Pascale Carayon Sara Kraemer. A human factors vulnerability evaluation method for computer and information security.

[6] Eugene Schultz. The human factor in security. *Computers & Security*, 24:425–426, 2005.

[7] Lesia L. Crumpton Pamela R. McCauley-Bell. The human factors issues in information security: What are they and do they matter? In *Proceedings of the Human Factors and Ergonomics Society*, pages 439–443, 1998.

[8] Erkan Kahraman. Evaluating it security performance with quantifiable metrics. Master's thesis, DSV SU/KTH, 2005.

[9] Gary Hinson. Human factors in information security.

[10] Agata Sawicka Jose J. Gonzalez. A framework for human factors in information security.

[11] Bruce Schneier. *Secrets and Lies*. Robert Ipsen, 2000.

[12] William L. Simon Kevin D. Mitnik. *The Art of Deception*. Wiley Publishing, Inc., 2002.

[13] Konstantin Sapronov. The human factor and information security, December 2005.

[14] Dancho Danchev. Reducing "human factor" mistakes, March 2006.

[15] Raymond R. Panko. *Corporate Computer and Network Security*. Pearson Education, Inc., 2004.

[16] Terry Timm Moos. Cisco-sponsored security survey of remote workers reveals the need for more user awareness, November 2006.

[17] Eric Maiwald. *Fundamentals of Network Security*. McGraw-Hill Technology Education, 2004.

[18] Ivn Arce. The weakest link revisited. *IEEE Security and Privacy*, 1:72–76, 2003.

[19] Michele D. Crabb David L. Oppenheimer, David A. Wagner. System security: A management perspective, September 1997.

[20] Robert E. Newman. *Enterprise Security*. Pearson Education, Inc., first edition, 2003.

[21] Neil J. Salkind. *Exploring Research*. Pearson Education, Inc., fifth edition, 2003.

[22] Alison Jane Pickard. *Research Methods in Information*. Facet Publishing, 2007.

[23] John W. Creswell. *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*. SAGE Publications, Inc, third edition, 2008.

[24] Mike Speciner Charlie Kaufman, Radia Perlman. *Network Security: Private Communication in a Public World*. Prentice Hall, Inc., 1995.

[25] Axel Boldt. Bliss, a linux "virus", January 2000.

[26] Mikko Hypponen. F-secure computer virus information pages: Linux/staog, February 1997.

[27] Ero Carrera. F-secure computer virus information pages: Linux.devnull, September 2002.

[28] Joel R. Voss. Ssh bruteforce virus by altsci, December 2007.

[29] Ryan Naraine. Mac os x malware found in pirated apple iwork 09, January 2009.

[30] Ryan Naraine. ibotnet: Researchers find signs of zombie macs, April 2009.

[31] Jane Wakefield. Office intruder "steals" data, May 2009.

[32] Ariel Rabkin. Personal knowledge questions for fallback authentication: Security questions in the era of facebook. In *Proceedings of the 4th symposium on Usable privacy and security*, pages 13–23, 2008.

[33] Inc. Cisco Systems. Actions speak louder than words: Despite claiming security awareness, many remote workers engage in risky online behavior, October 2006.

[34] Jeremy L. Smith. I.t. security: Antivirus strategy - improving the security of the dispatch center. Technical report, Plant Equipment, Inc., 2005.

[35] Al Gillen John F. Gantz, Christian A. Christiansen. The risks of obtaining and using pirated software, October 2006.